JC17 Rec'd PCT/PTO 10/537121 02 JUN 2005

COMPOSITIONS, SPLICE VARIANTS AND METHODS RELATING TO COLON SPECIFIC GENES AND PROTEINS

INTRODUCTION

This application claims the benefit of priority from U.S. Provisional Patent Application Serial No. 60/431,143 filed December 4, 2002 and 60/431,206 filed December 4, 2002, which are herein incorporated by reference in their entireties.

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic colon cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, post translational modifications (PTMs), variants, derivatives, agonists and antagonists thereof and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating colon cancer and/or non-cancerous disease states in colon, identifying colon tissue and monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, therapeutic molecules including but not limited to antibodies or antisense molecules, production of transgenic animals and cells, and production of engineered colon tissue for treatment and research.

BACKGROUND OF THE INVENTION

Colorectal cancer is the second most common cause of cancer death in the United States and the third most prevalent cancer in both men and women. M. L. Davila & A. D. Davila, *Screening for Colon and Rectal Cancer, in* Colon and Rectal Cancer 47 (Peter S. Edelstein ed., 2000). The American Cancer Society estimates that there will be about 105,500 new cases of colon cancer and 42,000 new cases of rectal cancer in 2003 in the United States. Colon cancer and rectal cancer will cause about 57,100 deaths combined. ACS Website: cancer.org on the world wide web. Nearly all cases of colorectal cancer arise from adenomatous polyps, some of which mature into large polyps, undergo abnormal growth and development, and ultimately progress into cancer. Davila at 55-56. This progression would appear to take at least 10 years in most patients, rendering it a

readily treatable form of cancer if diagnosed early, when the cancer is localized. Davila at 56; Walter J. Burdette, Cancer: Etiology, Diagnosis, and Treatment 125 (1998).

Although our understanding of the etiology of colon cancer is undergoing continual refinement, extensive research in this area points to a combination of factors, including age, hereditary and nonhereditary conditions, and environmental/dietary factors. Age is a key risk factor in the development of colorectal cancer, Davila at 48, with men and women over 40 years of age becoming increasingly susceptible to that cancer, Burdette at 126. Incidence rates increase considerably in each subsequent decade of life. Davila at 48. A number of hereditary and nonhereditary conditions have also been linked to a heightened risk of developing colorectal cancer, including familial adenomatous polyposis (FAP), hereditary nonpolyposis colorectal cancer (Lynch syndrome or HNPCC), a personal and/or family history of colorectal cancer or adenomatous polyps, inflammatory bowel disease, diabetes mellitus, and obesity. Davila at 47; Henry T. Lynch & Jane F. Lynch, *Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndromes), in* Colon and Rectal Cancer 67-68 (Peter S. Edelstein ed., 2000).

Environmental/dietary factors associated with an increased risk of colorectal cancer include a high fat diet, intake of high dietary red meat, and sedentary lifestyle. Davila at 47; Reddy, B. S., *Prev. Med.* 16(4): 460-7 (1987). Conversely, environmental/dietary factors associated with a reduced risk of colorectal cancer include a diet high in fiber, folic acid, calcium, and hormone-replacement therapy in post-menopausal women. Davila at 50-55. The effect of antioxidants in reducing the risk of colon cancer is unclear. Davila at 53.

Because colon cancer is highly treatable when detected at an early, localized stage, screening should be a part of routine care for all adults starting at age 50, especially those with first-degree relatives with colorectal cancer. One major advantage of colorectal cancer screening over its counterparts in other types of cancer is its ability to not only detect precancerous lesions, but to remove them as well. Davila at 56. The key colorectal cancer screening tests in use today are fecal occult blood test, sigmoidoscopy, colonoscopy, double-contrast barium enema, and the carcinoembryonic antigen (CEA) test. Burdette at 125; Davila at 56.

The fecal occult blood test (FOBT) screens for colorectal cancer by detecting the amount of blood in the stool, the premise being that neoplastic tissue, particularly malignant tissue, bleeds more than typical mucosa, with the amount of bleeding increasing

with polyp size and cancer stage. Davila at 56-57. While effective at detecting early stage tumors, FOBT is unable to detect adenomatous polyps (premalignant lesions), and, depending on the contents of the fecal sample, is subject to rendering false positives. Davila at 56-59. Sigmoidoscopy and colonoscopy, by contrast, allow direct visualization of the bowel, and enable one to detect, biopsy, and remove adenomatous polyps. Davila at 59-60, 61. Despite the advantages of these procedures, there are accompanying downsides: sigmoidoscopy, by definition, is limited to the sigmoid colon and below, colonoscopy is a relatively expensive procedure, and both share the risk of possible bowel perforation and hemorrhaging. Davila at 59-60. Double-contrast barium enema (DCBE) enables detection of lesions better than FOBT, and almost as well a colonoscopy, but it may be limited in evaluating the winding rectosigmoid region. Davila at 60. The CEA blood test, which involves screening the blood for carcinoembryonic antigen, shares the downside of FOBT, in that it is of limited utility in detecting colorectal cancer at an early stage. Burdette at 125.

Once colon cancer has been diagnosed, treatment decisions are typically made in reference to the stage of cancer progression. A number of techniques are employed to stage the cancer (some of which are also used to screen for colon cancer), including pathologic examination of resected colon, sigmoidoscopy, colonoscopy, and various imaging techniques. <u>AJCC Cancer Staging Handbook</u> 84 (Irvin D. Fleming et al. eds., 5th ed. 1998); Montgomery, R. C. and Ridge, J.A., *Semin. Surg. Oncol.* 15(3): 143-150 (1998). Moreover, chest films, liver functionality tests, and liver scans are employed to determine the extent of metastasis. Fleming at 84. While computerized tomography and magnetic resonance imaging are useful in staging colorectal cancer in its later stages, both have unacceptably low staging accuracy for identifying early stages of the disease, due to the difficulty that both methods have in (1) revealing the depth of bowel wall tumor infiltration and (2) diagnosing malignant adenopathy. Thoeni, R. F., *Radiol. Clin. N. Am.* 35(2): 457-85 (1997). Rather, techniques such as transrectal ultrasound (TRUS) are preferred in this context, although this technique is inaccurate with respect to detecting small lymph nodes that may contain metastases. David Blumberg & Frank G. Opelka, *Neoadjuvant and Adjuvant Therapy for Adenocarcinoma of the Rectum, in* <u>Colon and Rectal Cancer</u> 316 (Peter S. Edelstein ed., 2000).

Several classification systems have been devised to stage the extent of colorectal cancer, including the Dukes' system and the more detailed International Union against

Cancer-American Joint Committee on Cancer TNM staging system, which is considered by many in the field to be a more useful staging system. Burdette at 126-27. The TNM system, which is used for either clinical or pathological staging, is divided into four stages, each of which evaluates the extent of cancer growth with respect to primary tumor (T), regional lymph nodes (N), and distant metastasis (M). Fleming at 84-85. The system focuses on the extent of tumor invasion into the intestinal wall, invasion of adjacent structures, the number of regional lymph nodes that have been affected, and whether distant metastasis has occurred. Fleming at 81.

Stage 0 is characterized by *in situ* carcinoma (Tis), in which the cancer cells are located inside the glandular basement membrane (intraepithelial) or lamina propria (intramucosal). In this stage, the cancer has not spread to the regional lymph nodes (N0), and there is no distant metastasis (M0). In stage I, there is still no spread of the cancer to the regional lymph nodes and no distant metastasis, but the tumor has invaded the submucosa (T1) or has progressed further to invade the muscularis propria (T2). Stage II also involves no spread of the cancer to the regional lymph nodes and no distant metastasis, but the tumor has invaded the subserosa, or the nonperitonealized pericolic or perirectal tissues (T3), or has progressed to invade other organs or structures, and/or has perforated the visceral peritoneum (T4). Stage III is characterized by any of the T substages, no distant metastasis, and either metastasis in 1 to 3 regional lymph nodes (N1) or metastasis in four or more regional lymph nodes (N2). Lastly, stage IV involves any of the T or N substages, as well as distant metastasis. Fleming at 84-85; Burdette at 127.

Currently, pathological staging of colon cancer is preferable over clinical staging as pathological staging provides a more accurate prognosis. Pathological staging typically involves examination of the resected colon section, along with surgical examination of the abdominal cavity. Fleming at 84. Clinical staging would be a preferred method of staging were it at least as accurate as pathological staging, as it does not depend on the invasive procedures of its counterpart.

Turning to the treatment of colorectal cancer, surgical resection results in a cure for roughly 50% of patients. Irradiation is used both preoperatively and postoperatively in treating colorectal cancer. Chemotherapeutic agents, particularly 5-fluorouracil, are also powerful weapons in treating colorectal cancer. Other agents include irinotecan and floxuridine, cisplatin, levamisole, methotrexate, interferon-α, and leucovorin. Burdette at 125, 132-33. Nonetheless, thirty to forty percent of patients will develop a recurrence of

colon cancer following surgical resection, which in many patients is the ultimate cause of death. Wayne De Vos, *Follow-up After Treatment of Colon Cancer*, Colon and Rectal Cancer 225 (Peter S. Edelstein ed., 2000). Accordingly, colon cancer patients must be closely monitored to determine response to therapy and to detect persistent or recurrent disease and metastasis.

The next few paragraphs describe the some of molecular bases of colon cancer. In the case of FAP, the tumor suppressor gene APC (adenomatous polyposis coli), chromosomally located at 5q21, has been either inactivated or deleted by mutation. Alberts et al., Molecular Biology of the Cell 1288 (3d ed. 1994). The APC protein plays a role in a number of functions, including cell adhesion, apoptosis, and repression of the *c-myc* oncogene. N. R. Hall & R. D. Madoff, *Genetics and the Polyp-Cancer Sequence*, Colon and Rectal Cancer 8 (Peter S. Edelstein, ed., 2000). Of those patients with colorectal cancer who have normal APC genes, over 65% have such mutations in the cancer cells but not in other tissues. Alberts et al., *supra* at 1288. In the case of HPNCC, patients manifest abnormalities in the tumor suppressor gene HNPCC, but only about 15% of tumors contain the mutated gene. *Id.* A host of other genes have also been implicated in colorectal cancer, including the K-*ras*, N-*ras*, H-*ras* and c-*myc* oncogenes, and the tumor suppressor genes *DCC* (deleted in colon carcinoma) and *p53*. Hall & Madoff, at 8-9; Alberts et al., at 1288.

Abnormalities in Wg/Wnt signal transduction pathway are also associated with the development of colorectal carcinoma. Taipale, J. and Beachy, P.A. *Nature* 411: 349-354 (2001). Wnt1 is a secreted protein gene originally identified within mouse mammary cancers by its insertion into the mouse mammary tumor virus (MMTV) gene. The protein is homologous to the wingless (Wg) gene product of Drosophila, in which it functions as an important factor for the determination of dorsal-ventral segmentation and regulates the formation of fly imaginal discs. Wg/Wnt pathway controls cell proliferation, death and differentiation. Taipal (2001). There are at least 13 members in the Wnt family. These proteins have been found expressed mainly in the central nervous system (CNS) of vertebrates as well as other tissues such as mammary and intestine. The Wnt proteins are the ligands for a family of seven transmembrane domain receptors related to the Frizzled gene product in Drosophila. Binding Wnt to Frizzled stimulates the activity of the downstream target, Dishevelled, which in turn inactivates the glycogen synthesase kinase 3β (GSK3β). Taipal (2001). Usually active GSK3β will form a complex with the

adenomatous polyposis coli (APC) protein and phosphorylate another complex member, β-catenin. Once phosphorylated, β-catenin is directed to degradation through the ubiquitin pathway. When GSK3β or APC activity is down regulated, β-catenin is accumulated in the cytoplasm and binds to the T-cell factor or lymphocyte excitation factor (Tcf/Lef) family of transcriptional factors. Binding of β-catenin to Tcf releases the transcriptional repression and induces gene transcription. Among the genes regulated by β-catenin are a transcriptional repressor Engrailed, a transforming growth factor-β (TGF-β) family member Decapentaplegic, and the cytokine Hedgehog in Drosophila. β-Catenin is also involved in regulating cell adhesion by binding to α-catenin and E-cadherin. On the other hand, binding of β-catenin to these proteins controls the cytoplasmic β-catenin level and its complexing with TCF. Taipal (2001). Growth factor stimulation and activation of c-src or v-src also regulate β-catenin level by phosphorylation of α-catenin and its related protein, $p120^{cas}$. When phosphorylated, these proteins decrease their binding to E-cadherin and β-catenin resulting in the accumulation of cytoplasmic β-catenin. Reynolds, A.B. et al. *Mol. Cell Biol.* 14: 8333-8342 (1994). In colon cancer, c-src enzymatic activity has been shown to be increased to the level of v-src. Alternation of components in the Wg/Wnt pathway promotes colorectal carcinoma development. The best known modifications are to the APC gene. Nicola S et al. *Hum. Mol. Genet* 10:721-733 (2001). This germline mutation causes the appearance of hundreds to thousands of adenomatous polyps in the large bowel. It is the gene defect that accounts for the autosomally dominantly inherited FAP and related syndromes. The molecular alternations that occur in this pathway largely involve deletions of alleles of tumor-suppressor genes, such as APC, p53 and Deleted in Colorectal Cancer (DCC), combined with mutational activation of proto-oncogenes, especially c-Ki-ras. Aoki, T. et al. *Human Mutat.* 3: 342-346 (1994). All of these lead to genomic instability in colorectal cancers.

Another source of genomic instability in colorectal cancer is the defect of DNA mismatch repair (MMR) genes. Human homologues of the bacterial *mut*HLS complex (hMSH2, hMLH1, hPMS1, hPMS2 and hMSH6), which is involved in the DNA mismatch repair in bacteria, have been shown to cause the HNPCC (about 70-90% HNPCC) when mutated. Modrich, P. and Lahue, R. *Ann Rev. Biochem.* 65: 101-133 (1996); and Peltomäki, P. *Hum. Mol. Genet* 10: 735-740 (2001). The inactivation of these proteins leads to the accumulation of mutations and causes genetic instability that represents errors

in the accurate replication of the repetitive mono-, di-, tri- and tetra-nucleotide repeats, which are scattered throughout the genome (microsatellite regions). Jass, J.R. et al. *J. Gastroenterol Hepatol* 17: 17-26 (2002). Like in the classic FAP, mutational activation of c-Ki-ras is also required for the promotion of MSI in the alternative HNPCC. Mutations in other proteins such as the tumor suppressor protein phosphatase PTEN (Zhou, X.P. et al. *Hum. Mol. Genet* 11: 445-450 (2002)), BAX (Buttler, L.M. *Aus. N. Z. J. Surg.* 69: 88-94 (1999)), Caspase-5 (Planck, M. *Cancer Genet Cytogenet.* 134: 46-54 (2002)), TGFβ-RII (Fallik, D. et al. *Gastroenterol Clin Biol.* 24: 917-22 (2000)) and IGFII-R (Giovannucci E. *J. Nutr.* 131: 3109S-20S (2001)) have also been found in some colorectal tumors possibly as the cause of MMR defect.

Some tyrosine kinases have been shown up-regulated in colorectal tumor tissues or cell lines like HT29. Skoudy, A. et al. *Biochem J.* 317 (Pt 1): 279-84 (1996). Focal adhesion kinase (FAK) and its up-stream kinase c-src and c-yes in colonic epithelial cells may play an important role in the promotion of colorectal cancers through the extracellular matrix (ECM) and integrin-mediated signaling pathways. Jessup, J.M. et al., *The molecular biology of colorectal carcinoma, in:* The Molecular Basis of Human Cancer. 251-268 (Coleman W.B. and Tsongalis G.J. Eds. 2002). The formation of c-src/FAK complexes may coordinately deregulate VEGF expression and apoptosis inhibition. Recent evidences suggest that a specific signal-transduction pathway for cell survival that implicates integrin engagement leads to FAK activation and thus activates PI-3 kinase and akt. In turn, akt phosphorylates BAD and blocks apoptosis in epithelial cells. The activation of c-src in colon cancer may induce VEGF expression through the hypoxia pathway. Other genes that may be implicated in colorectal cancer include Cox enzymes (Ota, S. et al. *Aliment Pharmacol. Ther.* 16 (Suppl 2): 102-106 (2002)), estrogen (al-Azzawi, F. and Wahab, M. *Climacteric* 5: 3-14 (2002)), peroxisome proliferator-activated receptor-γ (PPAR-γ) (Gelman, L. et al. *Cell Mol. Life Sci.* 55: 932-943 (1999)), IGF-I (Giovannucci (2001)), thymine DNA glycosylase (TDG) (Hardeland, U. et al. *Prog. Nucleic Acid Res. Mol. Biol.* 68: 235-253 (2001)) and EGF (Mendelsohn, J. *Endocrine-Related Cancer* 8: 3-9 (2001)).

Gene deletion and mutation are not the only causes for development of colorectal cancers. Epigenetic silencing by DNA methylation also accounts for the loss of function of colorectal cancer suppressor genes. A strong association between MSI and CpG island methylation has been well characterized in sporadic colorectal cancers with high MSI but

not in those of hereditary origin. In one experiment, DNA methylation of MLH1, CDKN2A, MGMT, THBS1, RARB, APC, and p14ARF genes has been shown in 80%, 55%, 23%, 23%, 58%, 35%, and 50% of 40 sporadic colorectal cancers with high MSI respectively. Yamamoto, H. et al. *Genes Chromosomes Cancer* 33: 322-325 (2002); and Kim, K.M. et al. *Oncogene.* 12;21(35): 5441-9 (2002). Carcinogen metabolism enzymes such as GST, NAT, CYP and MTHFR are also associated with an increased or decreased colorectal cancer risk. Pistorius, S. et al. *Kongressbd Dtsch Ges Chir Kongr* 118: 820-824 (2001); and Potter, J.D. *J. Natl. Cancer Inst.* 91: 916-932 (1999).

From the foregoing, it is clear that procedures used for detecting, diagnosing, monitoring, staging, prognosticating, and preventing the recurrence of colorectal cancer are of critical importance to the outcome of the patient. Moreover, current procedures, while helpful in each of these analyses, are limited by their specificity, sensitivity, invasiveness, and/or their cost. As such, highly specific and sensitive procedures that would operate by way of detecting novel markers in cells, tissues, or bodily fluids, with minimal invasiveness and at a reasonable cost, would be highly desirable.

Accordingly, there is a great need for more sensitive and accurate methods for predicting whether a person is likely to develop colorectal cancer, for diagnosing colorectal cancer, for monitoring the progression of the disease, for staging the colorectal cancer, for determining whether the colorectal cancer has metastasized, and for imaging the colorectal cancer. Following accurate diagnosis, there is also a need for less invasive and more effective treatment of colorectal cancer.

Growth and metastasis of solid tumors are also dependent on angiogenesis. Folkman, J., 1986, *Cancer Research*, 46, 467-473; Folkman, J., 1989, *Journal of the National Cancer Institute*, 82, 4-6. It has been shown, for example, that tumors which enlarge to greater than 2 mm must obtain their own blood supply and do so by inducing the growth of new capillary blood vessels. Once these new blood vessels become embedded in the tumor, they provide a means for tumor cells to enter the circulation and metastasize to distant sites such as liver, lung or bone. Weidner, N., *et al.*, 1991, *The New England Journal of Medicine*, 324(1), 1-8.

Angiogenesis, defined as the growth or sprouting of new blood vessels from existing vessels, is a complex process that primarily occurs during embryonic development. The process is distinct from vasculogenesis, in that the new endothelial cells lining the vessel arise from proliferation of existing cells, rather than differentiating from

stem cells. The process is invasive and dependent upon proteolyisis of the extracellular matrix (ECM), migration of new endothelial cells, and synthesis of new matrix components. Angiogenesis occurs during embryogenic development of the circulatory system; however, in adult humans, angiogenesis only occurs as a response to a pathological condition (except during the reproductive cycle in women).

Under normal physiological conditions in adults, angiogenesis takes place only in very restricted situations such as hair growth and wounding healing. Auerbach, W. and Auerbach, R., 1994, *Pharmacol Ther.* 63(3):265-3 11; Ribatti et al.,1991, *Haematologica* 76(4):3 11-20; Risau, 1997, *Nature* 386(6626):67 1-4. Angiogenesis progresses by a stimulus which results in the formation of a migrating column of endothelial cells. Proteolytic activity is focused at the advancing tip of this "vascular sprout", which breaks down the ECM sufficiently to permit the column of cells to infiltrate and migrate. Behind the advancing front, the endothelial cells differentiate and begin to adhere to each other, thus forming a new basement membrane. The cells then cease proliferation and finally define a lumen for the new arteriole or capillary.

Unregulated angiogenesis has gradually been recognized to be responsible for a wide range of disorders, including, but not limited to, cancer, cardiovascular disease, rheumatoid arthritis, psoriasis and diabetic retinopathy. Folkman, 1995, *Nat Med* 1(1):27-31; Isner, 1999, *Circulation* 99(13): 1653-5; Koch, 1998, *Arthritis Rheum* 41(6):951-62; Walsh, 1999, *Rheumatology* (Oxford) 38(2):103-12; Ware and Simons, 1997, *Nat Med* 3(2): 158-64.

Of particular interest is the observation that angiogenesis is required by solid tumors for their growth and metastases. Folkman, 1986 *supra*; Folkman 1990, *J Natl. Cancer Inst.*, 82(1) 4-6; Folkman, 1992, *Semin Cancer Biol* 3(2):65-71; Zetter, 1998, *Annu Rev Med* 49:407-24. A tumor usually begins as a single aberrant cell which can proliferate only to a size of a few cubic millimeters due to the distance from available capillary beds, and it can stay `dormant` without further growth and dissemination for a long period of time. Some tumor cells then switch to the angiogenic phenotype to activate endothelial cells, which proliferate and mature into new capillary blood vessels. These newly formed blood vessels not only allow for continued growth of the primary tumor, but also for the dissemination and recolonization of metastatic tumor cells. The precise mechanisms that control the angiogenic switch is not well understood, but it is believed that

neovascularization of tumor mass results from the net balance of a multitude of angiogenesis stimulators and inhibitors Folkman, 1995, *supra*.

One of the most potent angiogenesis inhibitors is endostatin identified by O'Reilly and Folkman. O'Reilly et al., 1997, *Cell* 88(2):277-85; O'Reilly et al., 1994, *Cell* 79(2):3 15-28. Its discovery was based on the phenomenon that certain primary tumors can inhibit the growth of distant metastases. O'Reilly and Folkman hypothesized that a primary tumor initiates angiogenesis by generating angiogenic stimulators in excess of inhibitors. However, angiogenic inhibitors, by virtue of their longer half life in the circulation, reach the site of a secondary tumor in excess of the stimulators. The net result is the growth of primary tumor and inhibition of secondary tumor. Endostatin is one of a growing list of such angiogenesis inhibitors produced by primary tumors. It is a proteolytic fragment of a larger protein: endostatin is a 20 kDa fragment of collagen XVIII (amino acid H1132-K1315 in murine collagen XVIII). Endostatin has been shown to specifically inhibit endothelial cell proliferation in vitro and block angiogenesis in vivo. More importantly, administration of endostatin to tumor-bearing mice leads to significant tumor regression, and no toxicity or drug resistance has been observed even after multiple treatment cycles. Boehm et al., 1997, Nature 390(6658):404-407. The fact that endostatin targets genetically stable endothelial cells and inhibits a variety of solid tumors makes it a very attractive candidate for anticancer therapy. Fidler and Ellis, 1994, Cell 79(2):185-8; Gastl et al., 1997, Oncology 54(3):177-84; Hinsbergh et al., 1999, Ann Oncol 10 Suppl 4:60-3. In addition, angiogenesis inhibitors have been shown to be more effective when combined with radiation and chemotherapeutic agents. Klement, 2000, J. Clin Invest, 105(8) R15-24. Browder, 2000, Cancer Res. 6-(7) 1878-86, Arap et al., 1998, Science 279(5349):377-80; Mauceri et al., 1998, Nature 394(6690):287-91.

SUMMARY OF THE INVENTION

The present invention solves many needs in the art by providing nucleic acid molecules, polypeptides and antibodies thereto, variants and derivatives of the nucleic acids and polypeptides, and agonists and antagonists thereto that may be used to identify, diagnose, monitor, stage, image and treat colon cancer and/or non-cancerous disease states in colon; identify and monitor colon tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy,

methods for producing transgenic animals and cells, and methods for producing engineered colon tissue for treatment and research.

One aspect of the present invention relates to nucleic acid molecules that are specific to colon cells, colon tissue and/or the colon organ. These colon specific nucleic acids (CSNAs) may be a naturally occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally occurring nucleic acid molecule. If the CSNA is genomic DNA, then the CSNA is a colon specific gene (CSG). If the CSNA is RNA, then it is a colon specific transcript encoded by a CSG. Due to alternative splicing and transcriptional modification one CSG may encode for multiple colon specific RNAs. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to colon. More preferred is a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 113-259. In another preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1-112. For the CSNA sequences listed herein, DEX0449_001.nt.1 corresponds to SEQ ID NO: 1. For sequences with multiple splice variants, the parent sequence DEX0449_001.nt.1, will be followed by DEX0449_001.nt.2, etc. for each splice variant. The sequences off the corresponding peptides are listed as DEX0449_001.aa.1, etc. For the mapping of all of the nucleotides and peptides, see the table in the Example 1 section below.

This aspect of the present invention also relates to nucleic acid molecules that selectively hybridize or exhibit substantial sequence similarity to nucleic acid molecules encoding a Colon Specific Protein (CSP), or that selectively hybridize or exhibit substantial sequence similarity to a CSNA. In one embodiment of the present invention the nucleic acid molecule comprises an allelic variant of a nucleic acid molecule encoding a CSP, or an allelic variant of a CSNA. In another embodiment, the nucleic acid molecule comprises a part of a nucleic acid sequence that encodes a CSP or a part of a nucleic acid sequence of a CSNA.

In addition, this aspect of the present invention relates to a nucleic acid molecule further comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of a CSNA or the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of a CSP.

Another aspect of the present invention relates to vectors and/or host cells comprising a nucleic acid molecule of this invention. In a preferred embodiment, the

nucleic acid molecule of the vector and/or host cell encodes all or a fragment of a CSP. In another preferred embodiment, the nucleic acid molecule of the vector and/or host cell comprises all or a part of a CSNA. Vectors and host cells of the present invention are useful in the recombinant production of polypeptides, particularly CSPs of the present invention.

Another aspect of the present invention relates to polypeptides encoded by a nucleic acid molecule of this invention. The polypeptide may comprise either a fragment or a full-length protein. In a preferred embodiment, the polypeptide is a CSP. However, this aspect of the present invention also relates to mutant proteins (muteins) of CSPs, fusion proteins of which a portion is a CSP, and proteins and polypeptides encoded by allelic variants of a CSNA as provided herein.

A further aspect of the present invention is a novel splice variant which encodes an amino acid sequence that provides a novel region to be targeted for the generation of reagents that can be used in the detection and/or treatment of cancer. The novel amino acid sequence may lead to a unique protein structure, protein subcellular localization, biochemical processing or function. This information can be used to directly or indirectly facilitate the generation of additional or novel therapeutics or diagnostics. The nucleotide sequence in this novel splice variant can be used as a nucleic acid probe for the diagnosis and/or treatment of cancer.

Another aspect of the present invention relates to antibodies and other binders that specifically bind to a polypeptide of the instant invention. Accordingly antibodies or binders of the present invention specifically bind to CSPs, muteins, fusion proteins, and/or homologous proteins or polypeptides encoded by allelic variants of a CSNA as provided herein.

Another aspect of the present invention relates to agonists and antagonists of the nucleic acid molecules and polypeptides of this invention. The agonists and antagonists of the instant invention may be used to treat colon cancer and non-cancerous disease states in colon and to produce engineered colon tissue.

Another aspect of the present invention relates to methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. Such methods are useful in identifying, diagnosing, monitoring, staging, imaging and treating colon cancer and/or non-cancerous disease states in colon. Such methods are also useful

in identifying and/or monitoring colon tissue. In addition, measurement of levels of one or more of the nucleic acid molecules of this invention may be useful as a diagnostic as part of a panel in combination with known other markers, particularly those described in the colon cancer background section above.

Another aspect of the present invention relates to use of the nucleic acid molecules of this invention in gene therapy, for producing transgenic animals and cells, and for producing engineered colon tissue for treatment and research.

Another aspect of the present invention relates to methods for detecting polypeptides of this invention, preferably using antibodies thereto. Such methods are useful to identify, diagnose, monitor, stage, image and treat colon cancer and non-cancerous disease states in colon. In addition, measurement of levels of one or more of the polypeptides of this invention may be useful to identify, diagnose, monitor, stage, and/or image colon cancer in combination with known other markers, particularly those described in the colon cancer background section above. The polypeptides of the present invention can also be used to identify and/or monitor colon tissue, and to produce engineered colon tissue.

Yet another aspect of the present invention relates to a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences. In addition, the computer records regarding the nucleic acid and/or amino acid sequences and/or measurements of their levels may be used alone or in combination with other markers to diagnose colon related diseases.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well known and commonly used in

the art. The methods and techniques of the present invention are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999).

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally occurring and modified nucleotides linked together by naturally occurring and/or non-naturally occurring nucleotide linkages.

Nucleotides are represented by single letter symbols in nucleic acid molecule sequences. The following table lists symbols identifying nucleotides or groups of

nucleotides which may occupy the symbol position on a nucleic acid molecule. *See* Nomenclature Committee of the International Union of Biochemistry (NC-IUB), Nomenclature for incompletely specified bases in nucleic acid sequences, Recommendations 1984., *Eur J Biochem.* 150(1):1-5 (1985).

Symbol	Meaning	Group/Origin of Designation	Complementary Symbol
a	a	Adenine	t/u
g	g	Guanine	c
c	c	Cytosine	g
t	t	Thymine	a
u	u	Uracil	a
r	g or a	puRine	y
y	t/u or c	pYrimidine	r
m	a or c	aMino	k
k	g or t/u	Keto	m
s	g or c	Strong interactions 3H-bonds	w
w	a or t/u	Weak interactions 2H-bonds	s
b	g or c or t/u	not a	v
d	a or g or t/u	not c	h
h	a or c or t/u	not g	d
v	a or g or c	not t, not u	b
n	a or g or c or t/u, unknown, or other	aNy	n

The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus provides a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and U.S. Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single-or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including in vitro recombinant DNA-mediated techniques and by

expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

The term "naturally occurring nucleotide" referred to herein includes naturally occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotide linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phoshoraniladate, phosphoroamidate, and the like. *See e.g.*, LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), and U.S. Patent No. 5,151,510, the disclosure of which is hereby incorporated by reference in its entirety.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence. In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998)). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, double-stranded RNA (dsRNA) inhibition (RNAi), combination of triplex and antisense, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology" interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists between a first and second nucleic acid sequence when the first nucleic acid sequence or fragment thereof hybridizes to an antisense strand of the second nucleic acid, under selective hybridization conditions. Typically, selective hybridization will occur between the first nucleic acid sequence and an antisense strand of the second nucleic acid sequence when there is at least about 55% sequence identity between the first and second nucleic acid sequences— preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% — over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. *See* Sambrook (1989), *supra*, p. 9.51.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$T_m = 81.5°C + 16.6 (\log_{10}[Na^+]) + 0.41$ (fraction G + C) – 0.63 (% formamide) - (600/l) where l is the length of the hybrid in base pairs.

The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

$T_m = 79.8°C + 18.5 (\log_{10}[Na^+]) + 0.58$ (fraction G + C) + 11.8 (fraction G + C)^2 - 0.35 (% formamide) - (820/l).

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

$T_m = 79.8°C + 18.5(\log_{10}[Na^+]) + 0.58$ (fraction G + C) + 11.8 (fraction G + C)^2 - 0.50 (% formamide) - (820/l).

In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well known in the art.

An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (*e.g.* 42°C and 6X SSC) and varying the formamide concentration from 50% to

0%. Hybridization buffers may also include blocking agents to lower background. These agents are well known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acids that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid is created synthetically or recombinantly using a high codon degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (*e.g.*, for oligonucleotide probes) may be calculated by the formula:

$T_m = 81.5°C + 16.6(\log_{10}[Na^+]) + 0.41(\text{fraction } G+C) - (600/N)$, wherein N is change length and the $[Na^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well known in the art. *See, e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 μg of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 μl of reaction buffer. For the purpose of

isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAs. Techniques for ligation are well known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies. In another aspect, the invention is directed to single exon probes based on the CSNAs disclosed herein.

In one embodiment, the term "microarray" refers to a "nucleic acid microarray" having a substrate-bound plurality of nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Nucleic acid microarrays include all the

devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). Additionally, these nucleic acid microarrays include a substrate-bound plurality of nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al., Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000). Examples of nucleic acid microarrays may be found in U.S. Patent Nos. 6,391,623, 6,383,754, 6,383,749, 6,380,377, 6,379,897, 6,376,191, 6,372,431, 6,351,712 6,344,316, 6,316,193, 6,312,906, 6,309,828, 6,309,824, 6,306,643, 6,300,063, 6,287,850, 6,284,497, 6,284,465, 6,280,954, 6,262,216, 6,251,601, 6,245,518, 6,263,287, 6,251,601, 6,238,866, 6,228,575, 6,214,587, 6,203,989, 6,171,797, 6,103,474, 6,083,726, 6,054,274, 6,040,138, 6,083,726, 6,004,755, 6,001,309, 5,958,342, 5,952,180, 5,936,731, 5,843,655, 5,814,454, 5,837,196, 5,436,327, 5,412,087, and 5,405,783, the disclosures of which are incorporated herein by reference in their entireties.

In an alternative embodiment, a "microarray" may also refer to a "peptide microarray" or "protein microarray" having a substrate-bound collection or plurality of polypeptides, the binding to each of the plurality of bound polypeptides being separately detectable. Alternatively, the peptide microarray may have a plurality of binders, including but not limited to monoclonal antibodies, polyclonal antibodies, phage display binders, yeast 2 hybrid binders, and aptamers, which can specifically detect the binding of the polypeptides of this invention. The array may be based on autoantibody detection to the polypeptides of this invention, see Robinson *et al., Nature Medicine* 8(3):295-301 (2002). Examples of peptide arrays may be found in WO 02/31463, WO 02/25288, WO 01/94946, WO 01/88162, WO 01/68671, WO 01/57259, WO 00/61806, WO 00/54046, WO 00/47774, WO 99/40434, WO 99/39210, and WO 97/42507 and U.S. Patent Nos. 6,268,210, 5,766,960, and 5,143,854, the disclosures of which are incorporated herein by reference in their entireties.

In addition, determination of the levels of the CSNA or CSP may be made in a multiplex manner using techniques described in WO 02/29109, WO 02/24959, WO 01/83502, WO01/73113, WO 01/59432, WO 01/57269, and WO 99/67641, the disclosures of which are incorporated herein by reference in their entireties.

The term "mutant", "mutated", or "mutation" when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment of the present invention, the nucleic acid sequence is the wild type nucleic acid sequence encoding a CSP or is a CSNA. The nucleic acid sequence may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.*, Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g.*, Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type

parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. *See, e.g.*, Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. *See, e.g.*, Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993).

"Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is either contiguous with the gene of interest to control the gene of interest, or acts in *trans* or at a distance to control the gene of interest.

The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional

components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which a recombinant expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refers to that portion of a transcript-derived nucleic acid that can be translated in its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence is meant to be inclusive of all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally occurring and non-naturally occurring proteins and polypeptides, as well as polypeptide fragments and polypeptide mutants, derivatives and analogs thereof. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises a CSP encoded by a nucleic acid molecule of the instant invention, or a fragment, mutant, analog or derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be determined by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel with a stain well known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well known in the art for purification.

The term "fragment" when used herein with respect to polypeptides of the present invention refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length CSP. In a preferred embodiment, the fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally occurring polypeptide. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" when used herein with respect to polypeptides of the present invention refers to a polypeptide which is substantially similar in primary structural sequence to a CSP but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the CSP. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modifications include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, ^{14}C and ^{3}H, ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*.

The term "fusion protein" refers to polypeptides of the present invention coupled to a heterologous amino acid sequence. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence that encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --$CH(OH)CH_2$--, and --CH_2SO--, by methods well known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more

stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al., Ann. Rev. Biochem.* 61:387-418 (1992)). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

The term "mutant" or "mutein" when referring to a polypeptide of the present invention relates to an amino acid sequence containing substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a CSP. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to a CSP. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as GAP or BESTFIT or other variation Smith-Waterman alignment. *See*, T. F. Smith and M. S. Waterman, J. Mol. Biol. 147:195-197 (1981) and W.R. Pearson, Genomics 11:635-650 (1991).

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not

substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterize the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins. Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991).

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. *See* Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed., Sinauer Associates (1991). Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α-, α-disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ-carboxyglutamate, ε-N,N,N-trimethyllysine, ε-N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (*e.g.*, 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

By "homology" or "homologous" when referring to a polypeptide of the present invention it is meant polypeptides from different organisms with a similar sequence to the encoded amino acid sequence of a CSP and a similar biological activity or function. Although two polypeptides are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the polypeptides. Instead, the term "homologous" is defined to mean that the two polypeptides have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous polypeptide is one that exhibits 50% sequence similarity to CSP, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous polypeptides that exhibit 80%, 85% or 90% sequence similarity to a CSP. In yet a more preferred embodiment, a homologous polypeptide exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to polypeptides, it is recognized that residue positions that are not identical often differ by conservative amino acid

substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. *See, e.g.*, Pearson, *Methods Mol. Biol.* 24: 307-31 (1994).

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

1) Serine (S), Threonine (T);
2) Aspartic Acid (D), Glutamic Acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992). A "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. *See, e.g.*, GCG Version 6.1. Other programs include FASTA, discussed *supra*.

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. *See, e.g.,* Altschul *et al., J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al., Nucleic Acids Res.* 25:3389-402 (1997). Preferred parameters for blastp are:

Expectation value:	10 (default)
Filter:	seg (default)
Cost to open a gap:	11 (default)
Cost to extend a gap:	1 (default
Max. alignments:	100 (default)
Word size:	11 (default)
No. of descriptions:	100 (default)
Penalty Matrix:	BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

Algorithms other than blastp for database searching using amino acid sequences are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1.

An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', $F(ab')_2$, Fv, dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an

immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. A Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; a $F(ab')_2$ fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; a Fd fragment consists of the VH and CH1 domains; a Fv fragment consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g.*, Ward *et al.*, *Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" as used herein it is meant the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which VL and VH regions are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g.*, Bird *et al.*, *Science* 242: 423-426 (1988); Huston *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g.*, Holliger *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al.*, *Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specific binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than1 μM, preferably less than100 nM and most preferably less than 10 nM.

The term "patient" includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "colon specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the colon as compared to other tissues in the body. In a preferred embodiment, a "colon specific" nucleic acid molecule or polypeptide is detected at a level that is 1.5-fold higher than any other tissue in the body. In a more preferred embodiment, the "colon specific" nucleic acid molecule or polypeptide is detected at a level that is 2-fold higher than any other tissue in the body, more preferably 5-fold higher, still more preferably at least 10-fold, 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

One aspect of the invention provides isolated nucleic acid molecules that are specific to the colon or to colon cells or tissue or that are derived from such nucleic acid molecules. These isolated colon specific nucleic acids (CSNAs) may comprise cDNA genomic DNA, RNA, or a combination thereof, a fragment of one of these nucleic acids, or may be a non-naturally occurring nucleic acid molecule. A CSNA may be derived from an animal. In a preferred embodiment, the CSNA is derived from a human or other mammal. In a more preferred embodiment, the CSNA is derived from a human or other primate. In an even more preferred embodiment, the CSNA is derived from a human.

In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to colon, a colon-specific polypeptide (CSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 113-259. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1-112. Nucleotide sequences of the instantly-described nucleic acid molecules were determined by assembling several DNA molecules from either public or proprietary databases. Some of the underlying DNA sequences are the result, directly or indirectly, of at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Amersham Biosciences, Sunnyvale, CA, USA).

Nucleic acid molecules of the present invention may also comprise sequences that selectively hybridize to a nucleic acid molecule encoding a CSNA or a complement or antisense thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may or may not encode a CSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes a CSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 113-259. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1-112 or the antisense sequence thereof. Preferably, the nucleic acid molecule selectively hybridizes to

a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a CSP under low stringency conditions. More preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a CSP under moderate stringency conditions. Most preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a CSP under high stringency conditions. In a preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 113-259. In a more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1-112.

Nucleic acid molecules of the present invention may also comprise nucleic acid sequences that exhibit substantial sequence similarity to a nucleic acid encoding a CSP or a complement of the encoding nucleic acid molecule. In this embodiment, it is preferred that the nucleic acid molecule exhibit substantial sequence similarity to a nucleic acid molecule encoding human CSP. More preferred is a nucleic acid molecule exhibiting substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 113-259. By substantial sequence similarity it is meant a nucleic acid molecule having at least 60%, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85% sequence identity with a nucleic acid molecule encoding a CSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 113-259. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90%, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99% sequence identity with a nucleic acid molecule encoding a CSP. Most preferred in this embodiment is a nucleic acid molecule that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a CSP.

The nucleic acid molecules of the present invention are also inclusive of those exhibiting substantial sequence similarity to a CSNA or its complement. In this embodiment, it is preferred that the nucleic acid molecule exhibit substantial sequence similarity to a nucleic acid molecule having a nucleic acid sequence of SEQ ID NO: 1-

112. By substantial sequence similarity it is meant a nucleic acid molecule that has at least 60%, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85% sequence identity with a CSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1-112. More preferred is a nucleic acid molecule that has at least 90%, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99% sequence identity with a CSNA. Most preferred is a nucleic acid molecule that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a CSNA.

Nucleic acid molecules that exhibit substantial sequence similarity are inclusive of sequences that exhibit sequence identity over their entire length to a CSNA or to a nucleic acid molecule encoding a CSP, as well as sequences that are similar over only a part of its length. In this case, the part is at least 50 nucleotides of the CSNA or the nucleic acid molecule encoding a CSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 113-259 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1-112. The similar nucleic acid molecule may also be a naturally occurring nucleic acid molecule from a human, when the CSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed

mutation of a CSNA. In a preferred embodiment, the substantially similar nucleic acid molecule is a CSNA.

The nucleic acid molecules of the present invention are also inclusive of allelic variants of a CSNA or a nucleic acid encoding a CSP. For example, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes and the sequence determined from one individual of a species may differ from other allelic forms present within the population. More than 1.4 million SNPs have already been identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001) – Variants with small deletions and insertions of more than a single nucleotide are also found in the general population, and often do not alter the function of the protein. In addition, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into a mRNA that encodes a CSP. In a more preferred embodiment, the gene is transcribed into a mRNA that encodes a CSP comprising an amino acid sequence of SEQ ID NO: 113-259. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into a mRNA that is a CSNA. In a more preferred embodiment, the gene is transcribed into a mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1-112. Also preferred is that the allelic variant be a naturally occurring allelic variant in the species of interest, particularly human.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences comprising a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is a CSP. In a preferred embodiment, the part encodes a CSP. In one embodiment, the nucleic acid molecule comprises a part of a CSNA. In another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a CSNA. In another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that is an allelic variant of a CSNA. In yet another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that encodes a CSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences that encode fusion proteins, homologous proteins, polypeptide fragments, muteins and polypeptide analogs, as described *infra*.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences containing modifications of the native nucleic acid molecule. Examples of such modifications include, but are not limited to, nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that may be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

Accordingly, in one embodiment, a nucleic acid molecule may include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. The labeled nucleic acid molecules are particularly useful as hybridization probes.

Common radiolabeled analogues include those labeled with ^{33}P, ^{32}P, and ^{35}S, such as α-^{32}P-dATP, α-^{32}P-dCTP, α-^{32}P-dGTP, α-^{32}P-dTTP, α-^{32}P-3'dATP, α-^{32}P-ATP, α-^{32}P-CTP, α-^{32}P-GTP, α-^{32}P-UTP, α-^{35}S-dATP, γ-^{35}S-GTP, γ-^{33}P-dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Biosciences, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas

Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. *See* Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000).

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules of the present invention can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and Peptide Nucleic Acids (PNA) to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); *see* Alers *et al.*, *Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994); Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994). Alternatively, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g.,* Tyagi *et al., Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al., Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al., Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al., Science* 279: 1228-1229 (1998); Marras *et al., Genet. Anal.* 14: 151-156 (1999); Holland *et al., Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al., Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al., Nucleic Acids Symp. Ser.* (37): 255-6 (1997); and U.S. Patent Nos. 5,846,726, 5,925,517, 5,925,517, 5,723,591 and 5,538,848, the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the present invention may also be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. *See* Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.), Oligonucleotides as Therapeutic Agents – Symposium No. 209, John Wiley & Son Ltd (1997). Such altered internucleoside bonds are often desired for techniques or for targeted gene correction, Gamper *et al., Nucl. Acids Res.* 28(21): 4332-4339 (2000). For double-stranded RNA inhibition which may utilize either natural ds RNA or ds RNA modified in its, sugar, phosphate or base, *see* Hannon, *Nature* 418(11): 244-251 (2002); Fire *et al.* in WO 99/32619; Tuschl *et al.* in US2002/0086356; Kruetzer *et al.* in WO 00/44895, the disclosures of which are incorporated herein by reference in their entirety. For circular antisense, see Kool in U.S. Patent No. 5,426,180, the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity

wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative U.S. Patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U.S. Patent Nos. 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S. Patent Nos. 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred nucleic acid molecules, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, U.S. Patent Nos.

5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference in its entirety. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA). PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The Tm of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the Tm of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the Tm by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the Tm by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with nucleobase sidechains. *See, e.g.*, Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999).

Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), and U.S. Patent Nos. 5,760,012 and 5,731,181, the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acid molecules of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed,

quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlocked conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); and Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994). Triplexed and quadruplexed conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997); Rowley *et al.*, *Mol Med* 5(10): 693-700 (1999); Kool, *Annu Rev Biophys Biomol Struct.* 25: 1-28 (1996).

SNP Polymorphisms

Commonly, sequence differences between individuals involve differences in single nucleotide positions. SNPs may account for 90% of human DNA polymorphism. Collins *et al.*, 8 *Genome Res.* 1229-31 (1998). SNPs include single base pair positions in genomic DNA at which different sequence alternatives (alleles) exist in a population. In addition, the least frequent allele generally must occur at a frequency of 1% or greater. DNA sequence variants with a reasonably high population frequency are observed approximately every 1,000 nucleotide across the genome, with estimates as high as 1 SNP per 350 base pairs. Wang *et al.*, 280 *Science* 1077-82 (1998); Harding *et al.*, 60 Am. J. Human Genet. 772-89 (1997); Taillon-Miller *et al.*, 8 *Genome Res.* 748-54 (1998); Cargill *et al.*, 22 *Nat. Genet.* 231-38 (1999); and Semple *et al.*, 16 *Bioinform. Disc. Note* 735-38 (2000). The frequency of SNPs varies with the type and location of the change. In base substitutions, two-thirds of the substitutions involve the C-T and G-A type. This variation in frequency can be related to 5-methylcytosine deamination reactions that occur frequently, particularly at CpG dinucleotides. Regarding location, SNPs occur at a much higher frequency in non-coding regions than in coding regions. Information on over one million variable sequences is already publicly available via the Internet and more such markers are available from commercial providers of genetic information. Kwok and Gu, 5 *Med. Today* 538-53 (1999).

Several definitions of SNPs exist. See, e.g., Brooks, 235 *Gene* 177-86 (1999). As used herein, the term "single nucleotide polymorphism" or "SNP" includes all single base variants, thus including nucleotide insertions and deletions in addition to single nucleotide substitutions. There are two types of nucleotide substitutions. A transition is the

replacement of one purine by another purine or one pyrimidine by another pyrimidine. A transversion is the replacement of a purine for a pyrimidine, or vice versa.

Numerous methods exist for detecting SNPs within a nucleotide sequence. A review of many of these methods can be found in Landegren *et al.*, 8 *Genome Res.* 769-76 (1998). For example, a SNP in a genomic sample can be detected by preparing a Reduced Complexity Genome (RCG) from the genomic sample, then analyzing the RCG for the presence or absence of a SNP. See, e.g., WO 00/18960 which is herein incorporated by reference in its entirety. Multiple SNPs in a population of target polynucleotides in parallel can be detected using, for example, the methods of WO 00/50869 which is herein incorporated by reference in its entirety. Other SNP detection methods include the methods of U.S. Pat. Nos. 6,297,018 and 6,322,980 which are herein incorporated by reference in their entirety. Furthermore, SNPs can be detected by restriction fragment length polymorphism (RFLP) analysis. See, e.g., U.S. Pat. Nos. 5,324,631; 5,645,995 which are herein incorporated by reference in their entirety. RFLP analysis of SNPs, however, is limited to cases where the SNP either creates or destroys a restriction enzyme cleavage site. SNPs can also be detected by direct sequencing of the nucleotide sequence of interest. In addition, numerous assays based on hybridization have also been developed to detect SNPs and mismatch distinction by polymerases and ligases. Several web sites provide information about SNPs including Ensembl on the World Wide Web at ensemble.org, Sanger Institute on the World Wide Web at sanger.ac.uk/genetics/exon/, National Center for Biotechnology Information (NCBI) on the World Wide Web at ncbi.nlm.nih.gov/SNP/ , The SNP Consortium Ltd. on the World Wide Web at snp.cshl.org. The chromosomal locations for the compositions disclosed herein are provided below. In addition, one of ordinary skill in the art could use a BLAST against the genome or any of the databases cited above to find the chromosomal location. Another a preferred method to find the genomic coordinates and associated SNPs would be to use the BLAT tool (genome.ucsc.edu, Kent et al. 2001, The Human Genome Browser at UCSC, Genome Research 996-1006 or Kent 2002 BLAT —The BLAST -Like Alignment Tool Genome Reseach, 1-9). All web sites above were accessed December 3, 2003.

RNA interference

RNA interference refers to the process of sequence-specific post transcriptional gene silencing in animals mediated by short interfering RNAs (siRNA). Fire *et al.*, 1998, *Nature*, 391, 806. The corresponding process in plants is commonly referred to as post transcriptional gene silencing or RNA silencing and is also referred to as quelling in fungi. The process of post transcriptional gene silencing is thought to be an evolutionarily conserved cellular defense mechanism used to prevent the expression of foreign genes which is commonly shared by diverse flora and phyla. Fire *et al.*, 1999, *Trends Genet.*, 15, 358. Such protection from foreign gene expression may have evolved in response to the production of double-stranded RNAs (dsRNA) derived from viral infection or the random integration of transposon elements into a host genome via a cellular response that specifically destroys homologous single-stranded RNA or viral genomic RNA. The presence of dsRNA in cells triggers the RNAi response though a mechanism that has yet to be fully characterized. This mechanism appears to be different from the interferon response that results from dsRNA mediated activation of protein kinase PKR and 2',5'-oligoadenylate synthetase resulting in non-specific cleavage of mRNA by ribonuclease L.

The presence of long dsRNAs in cells stimulates the activity of a ribonuclease III enzyme referred to as dicer. Dicer is involved in the processing of the dsRNA into short pieces of dsRNA known as short interfering RNAs (siRNA). Berstein *et al.*, 2001, *Nature*, 409, 363. Short interfering RNAs derived from dicer activity are typically about 21-23 nucleotides in length and comprise about 19 base pair duplexes. Dicer has also been implicated in the excision of 21 and 22 nucleotide small temporal RNAs (stRNA) from precursor RNA of conserved structure that are implicated in translational control. Hutvagner *et al.*, 2001, *Science*, 293, 834. The RNAi response also features an endonuclease complex containing a siRNA, commonly referred to as an RNA-induced silencing complex (RISC), which mediates cleavage of single-stranded RNA having sequence complementary to the antisense strand of the siRNA duplex. Cleavage of the target RNA takes place in the middle of the region complementary to the antisense strand of the siRNA duplex. Elbashir *et al.*, 2001, *Genes Dev.*, 15, 188.

Short interfering RNA mediated RNAi has been studied in a variety of systems. Fire *et al.*, 1998, *Nature*, 391, 806, were the first to observe RNAi in C. Elegans. Wianny and Goetz, 1999, *Nature Cell Biol.*, 2, 70, describe RNAi mediated by dsRNA in mouse embryos. Hammond *et al.*, 2000, *Nature*, 404, 293, describe RNAi in Drosophila cells

transfected with dsRNA. Elbashir *et al.*, 2001, *Nature*, 411, 494, describe RNAi induced by introduction of duplexes of synthetic 21-nucleotide RNAs in cultured mammalian cells including human embryonic kidney and HeLa cells. Recent work in Drosophila embryonic lysates (Elbashir *et al.*, 2001, *EMBO J.*, 20, 6877) has revealed certain requirements for siRNA length, structure, chemical composition, and sequence that are essential to mediate efficient RNAi activity. These studies have shown that 21 nucleotide siRNA duplexes are most active when containing two nucleotide 3'-overhangs. Furthermore, complete substitution of one or both siRNA strands with 2'-deoxy (2'-H) or 2'-O-methyl nucleotides abolishes RNAi activity, whereas substitution of the 3'-terminal siRNA overhang nucleotides with deoxy nucleotides (2'-H) was shown to be tolerated. Single mismatch sequences in the center of the siRNA duplex were also shown to abolish RNAi activity. In addition, these studies also indicate that the position of the cleavage site in the target RNA is defined by the 5'-end of the siRNA guide sequence rather than the 3'-end. Elbashir et al., 2001, *EMBO J.*, 20, 6877. Other studies have indicated that a 5'-phosphate on the target-complementary strand of a siRNA duplex is required for siRNA activity and that ATP is utilized to maintain the 5'-phosphate moiety on the siRNA. Nykanen et al., 2001, *Cell*, 107, 309.

Studies have shown that replacing the 3'-overhanging segments of a 21-mer siRNA duplex having 2 nucleotide 3' overhangs with deoxyribonucleotides does not have an adverse effect on RNAi activity. Replacing up to 4 nucleotides on each end of the siRNA with deoxyribonucleotides has been reported to be well tolerated whereas complete substitution with deoxyribonucleotides results in no RNAi activity. Elbashir *et al.*, 2001, *EMBO J.*, 20, 6877. In addition, Elbashir et al., supra, also report that substitution of siRNA with 2'-O-methyl nucleotides completely abolishes RNAi activity. Li et al., WO 00/44914, and Beach et al., WO 01/68836 both suggest that siRNA "may include modifications to either the phosphate-sugar back bone or the nucleoside to include at least one of a nitrogen or sulfur heteroatom", however neither application teaches to what extent these modifications are tolerated in siRNA molecules nor provide any examples of such modified siRNA. Kreutzer and Limmer, Canadian Patent Application No. 2,359,180, also describe certain chemical modifications for use in dsRNA constructs in order to counteract activation of double-stranded RNA-dependent protein kinase PKR, specifically 2'-amino or 2'-O-methyl nucleotides, and nucleotides containing a 2'-O or 4'-C methylene bridge. However, Kreutzer and Limmer similarly fail to show to what extent these modifications

are tolerated in siRNA molecules nor do they provide any examples of such modified siRNA.

Parrish et al., 2000, *Molecular Cell*, 6, 1977-1087, tested certain chemical modifications targeting the unc-22 gene in C. elegans using long (>25 nt) siRNA transcripts. The authors describe the introduction of thiophosphate residues into these siRNA transcripts by incorporating thiophosphate nucleotide analogs with T7 and T3 RNA polymerase and observed that "RNAs with two [phosphorothioate] modified bases also had substantial decreases in effectiveness as RNAi triggers; [phosphorothioate] modification of more than two residues greatly destabilized the RNAs in vitro and we were not able to assay interference activities." Parrish et al. at 1081. The authors also tested certain modifications at the 2'-position of the nucleotide sugar in the long siRNA transcripts and observed that substituting deoxynucleotides for ribonucleotides "produced a substantial decrease in interference activity", especially in the case of Uridine to Thymidine and/or Cytidine to deoxy-Cytidine substitutions. Parrish et al. In addition, the authors tested certain base modifications, including substituting 4-thiouracil, 5-bromouracil, 5-iodouracil, 3-(aminoallyl)uracil for uracil, and inosine for guanosine in sense and antisense strands of the siRNA, and found that whereas 4-thiouracil and 5-bromouracil were all well tolerated, inosine "produced a substantial decrease in interference activity" when incorporated in either strand. Incorporation of 5-iodouracil and 3-(aminoallyl)uracil in the antisense strand resulted in substantial decrease in RNAi activity as well.

Beach et al., WO 01/68836, describes specific methods for attenuating gene expression using endogenously derived dsRNA. Tuschl et al., WO 01/75164, describes a Drosophila in vitro RNAi system and the use of specific siRNA molecules for certain functional genomic and certain therapeutic applications; although Tuschl, 2001, *Chem. Biochem.*, 2, 239-245, doubts that RNAi can be used to cure genetic diseases or viral infection due "to the danger of activating interferon response". Li et al., WO 00/44914, describes the use of specific dsRNAs for use in attenuating the expression of certain target genes. Zernicka-Goetz et al., WO 01/36646, describes certain methods for inhibiting the expression of particular genes in mammalian cells using certain dsRNA molecules. Fire et al., WO 99/32619, U.S. Patent No. 6,506,559, the contents of which are hereby incorporated by reference in their entirety, describes particular methods for introducing

certain dsRNA molecules into cells for use in inhibiting gene expression. Plaetinck et al., WO 00/01846, describes certain methods for identifying specific genes responsible for conferring a particular phenotype in a cell using specific dsRNA molecules. Mello et al., WO 01/29058, describes the identification of specific genes involved in dsRNA mediated RNAi. Deschamps Depaillette et al., International PCT Publication No. WO 99/07409, describes specific compositions consisting of particular dsRNA molecules combined with certain anti-viral agents. Driscoll et al., International PCT Publication No. WO 01/49844, describes specific DNA constructs for use in facilitating gene silencing in targeted organisms. Parrish et al., 2000, Molecular Cell, 6, 1977-1087, describes specific chemically modified siRNA constructs targeting the unc-22 gene of C. elegans. Tuschl et al., International PCT Publication No. WO 02/44321, describe certain synthetic siRNA constructs.

Methods for Using Nucleic Acid Molecules as Probes and Primers

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect and characterize gross alterations in the gene of a CSNA, such as deletions, insertions, translocations, and duplications of the CSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g.*, Andreeff *et al.* (eds.), Introduction to Fluorescence *In Situ* Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999). The isolated nucleic acid molecules of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include a nucleic acid molecule of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

Alternatively, detection techniques such as molecular beacons may be used, see Kostrikis *et al. Science* 279:1228-1229 (1998).

The isolated nucleic acid molecules of the present invention can also be used as probes to detect, characterize, and quantify CSNA in, and isolate CSNA from, transcript-derived nucleic acid samples. In one embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A^{+}- selected RNA samples. In another embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g.*, Schwarchzacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000). In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to CSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000).

In another embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify and/or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In this embodiment, it is preferred that the probe or primer be derived from a nucleic acid molecule encoding a CSP. More preferably, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 113-259. Also preferred are probes or primers derived from a CSNA. More preferred are probes or primers derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-112.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides

in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); and McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995). Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; and Siebert (ed.), PCR Technique:RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995).

PCR and hybridization methods may be used to identify and/or isolate nucleic acid molecules of the present invention including allelic variants, homologous nucleic acid molecules and fragments. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules of the present invention that encode homologous proteins, analogs, fusion proteins or muteins of the invention. Nucleic acid primers as described herein can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as the template.

These nucleic acid primers can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Pat. No. 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7

(2001); International Patent publications WO 97/19193 and WO 00/15779, and U.S. Patent Nos. 5,854,033 and 5,714,320, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al., Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect

of the invention to provide microarrays that comprise one or more of the nucleic acid molecules of the present invention.

In yet another embodiment, the invention is directed to single exon probes based on the CSNAs disclosed herein.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention provides vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acid molecules of the present invention in host cells (cloning vectors), for shuttling the nucleic acid molecules of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acid molecules of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acid molecules of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acid molecules of the present invention, alone or as fusion proteins with heterologous polypeptides (expression vectors). Vectors are by now well known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*. Furthermore, a variety of vectors are available commercially. Use of existing vectors and modifications thereof are well within the skill in the art. Thus, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences that control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic acid sequence of this invention to an expression control sequence, of course, includes, if not already part of

the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λGT10 and λGT11, and other phages, *e.g.*, M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74:

527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as ura3-52, his3-D1, leu2-D1, trp1-D1 and lys2-201.

Insect cells may be chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda, e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

The host cells may also be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include, but are not limited to, resistance to neomycin (G418), blasticidin, hygromycin and zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and

retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the nucleic acid molecules of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the trc promoter, a hybrid derived from the trp and lac promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, and the araBAD operon. Prokaryotic expression vectors may further include transcription terminators, such as the aspA terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the CYC1 promoter, the GAL1 promoter, the GAL10 promoter, ADH1 promoter, the promoters of the yeast α-mating system, or the GPD promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the CYC1 or ADH1 gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include, but are not limited to, those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 and the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the CSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β-globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows a high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the trc promoter, which is regulated

by the lac operon, and the pL promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid Plac/ara-1 promoter and the PLtetO-1 promoter. The PLtetO-1 promoter takes advantage of the high expression levels from the PL promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the Tn10 tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one embodiment of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Such tags include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverley, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the polypeptides of the present invention can be expressed as a fusion to glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5

antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope, detectable by anti-HA antibody.

For secretion of expressed polypeptides, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful protein fusions include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusions for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the α-agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538

(AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well known in the art. *See* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999). A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g,* Cormack *et al.*, *Gene* 173: 33–38 (1996); U.S. Patent Nos. 6,090,919 and 5,804,387, the disclosures of which are incorporated herein by reference in their entireties) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g,.* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.*, Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.*, Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patent Nos. 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. *See* also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999); Yang, *et al.*, *J Biol Chem*, 273: 8212-6 (1998); Bevis *et al.*, *Nature Biotechnology*, 20:83-7 (2002). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half-life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor

and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, and WO 96/18412, the disclosures of which are incorporated herein by reference in their entireties.

For long-term, high-yield recombinant production of the polypeptides of the present invention, stable expression is preferred. Stable expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA) allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid molecules of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as an antibiotic or other selection marker, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed polypeptide in the desired fashion. Such

post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation, and it is an aspect of the present invention to provide CSPs with such post-translational modifications.

In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid molecules of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid molecules of this invention.

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid molecules according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell (*e.g.*, conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell,

vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae, Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda, e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (*e.g.*, from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well known and readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from colon are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human colon cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. *See, e.g.*, Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*.

Methods for introducing the vectors and nucleic acid molecules of the present invention into the host cells are well known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli,* in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (*e.g.*, Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli.*

Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, *e.g.*, with $CaCl_2$, or a solution of Mg^{2+}, Mn^{2+}, Ca^{2+}, Rb^+ or K^+, dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (*e.g.*, Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5α competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent E. coli Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent to take up exogenous DNA by electroporation by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided by BioRad (Richmond, CA, USA).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes such as a snail-gut extract, usually denoted Glusulase or Zymolyase, or an enzyme from *Arthrobacter luteus* to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+}. Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate to permeabilize the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded

carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with $CaPO_4$ or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for $CaPO_4$ transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent, FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found in, for example, ; Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000). Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.*, Cheng *et al.*, *Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

Purification of recombinantly expressed proteins is now well within the skill in the art and thus need not be detailed here. *See, e.g.*, Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and

Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001).

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Polypeptides, including Fragments Muteins, Homologous Proteins, Allelic Variants, Analogs and Derivatives

Another aspect of the invention relates to polypeptides encoded by the nucleic acid molecules described herein. In a preferred embodiment, the polypeptide is a colon specific polypeptide (CSP). In an even more preferred embodiment, the polypeptide comprises an amino acid sequence of SEQ ID NO:113-259 or is derived from a polypeptide having the amino acid sequence of SEQ ID NO: 113-259. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well known to those having ordinary skill in the art.

Polypeptides of the present invention may also comprise a part or fragment of a CSP. In a preferred embodiment, the fragment is derived from a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO: 113-259. Polypeptides of the present invention comprising a part or fragment of an entire CSP may or may not be CSPs. For example, a full-length polypeptide may be colon-specific, while a fragment thereof may be found in other tissues as well as in colon. A polypeptide that is not a CSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-CSP antibodies. In a preferred embodiment, the part or fragment is a CSP. Methods of determining whether a polypeptide of the present invention is a CSP are described *infra*.

Polypeptides of the present invention comprising fragments of at least 6 contiguous amino acids are also useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g.*, Geysen *et al., Proc. Natl. Acad. Sci. USA* 81: 3998-4002

(1984) and U.S. Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of a polypeptide of the present invention have utility in such a study.

Polypeptides of the present invention comprising fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize polypeptides of the present invention. *See, e.g.*, Lerner, Nature 299: 592-596 (1982); Shinnick *et al.*, *Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al.*, *Science* 219: 660-6 (1983). As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic and are capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the polypeptides of the present invention have utility as immunogens.

Polypeptides comprising fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire polypeptide, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the polypeptide of interest. See U.S. Patent Nos. 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

The polypeptide of the present invention thus preferably is at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the polypeptide of the present invention is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger polypeptides having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments by truncating the nucleic acid molecule, *e.g.*, a CSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally occurring polypeptide. Methods of producing polypeptide fragments are well known in the art. *See, e.g.*,

Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a fragment, preferably a fragment of a CSP, may be produced by chemical or enzymatic cleavage of a CSP polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule of the present invention encoding a fragment, preferably of a CSP, in a host cell.

Polypeptides of the present invention are also inclusive of mutants, fusion proteins, homologous proteins and allelic variants.

A mutant protein, or mutein, may have the same or different properties compared to a naturally occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native polypeptide. Small deletions and insertions can often be found that do not alter the function of a protein. Muteins may or may not be colon-specific. Preferably, the mutein is colon-specific. More preferably the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 113-259. Accordingly, in a preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 113-259. In a yet more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 113-259.

A mutein may be produced by isolation from a naturally occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein is produced from a host cell comprising a mutated nucleic acid molecule compared to the naturally occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid molecule of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino acids are altered, or may be

untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is colon-specific, as described below. Multiple random mutations can be introduced into the gene by methods well known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well known in the art. *See, e.g.*, Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), as well as U.S. Patent No. 5,223,408, which is herein incorporated by reference in its entirety.

The invention also contemplates polypeptides that are homologous to a polypeptide of the invention. In a preferred embodiment, the polypeptide is homologous to a CSP. In an even more preferred embodiment, the polypeptide is homologous to a CSP selected from the group having an amino acid sequence of SEQ ID NO: 113-259. By homologous polypeptide it is meant one that exhibits significant sequence identity to a CSP, preferably a CSP having an amino acid sequence of SEQ ID NO: 113-259. By significant sequence identity it is meant that the homologous polypeptide exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 113-259. More preferred are homologous polypeptides exhibiting at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 113-259. Most preferably, the homologous polypeptide exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 113-259. In a preferred embodiment, the amino acid substitutions of the homologous polypeptide are conservative amino acid substitutions as discussed supra.

Homologous polypeptides of the present invention also comprise polypeptide encoded by a nucleic acid molecule that selectively hybridizes to a CSNA or an antisense sequence thereof. In this embodiment, it is preferred that the homologous polypeptide be encoded by a nucleic acid molecule that hybridizes to a CSNA under low stringency,

moderate stringency or high stringency conditions, as defined herein. More preferred is a homologous polypeptide encoded by a nucleic acid sequence which hybridizes to a CSNA selected from the group consisting of SEQ ID NO: 1-112 or a homologous polypeptide encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes a CSP, preferably a CSP of SEQ ID NO:113-259 under low stringency, moderate stringency or high stringency conditions, as defined herein.

Homologous polypeptides of the present invention may be naturally occurring and derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, or baboon, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 113-259. The homologous polypeptide may also be a naturally occurring polypeptide from a human, when the CSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. The homologous polypeptide may also be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. Alternatively, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a CSP. In a preferred embodiment, the homologous polypeptide encodes a polypeptide that is a CSP.

Relatedness of proteins can also be characterized using a second functional test, such as the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated polpeptides not only identical in sequence to those described with particularity herein, but also to provide isolated polypeptides ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, polypeptides of the present invention are also inclusive of those encoded by an allelic variant of a nucleic acid molecule encoding a CSP. In this embodiment, it is preferred that the polypeptide be encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 113-259. More preferred is that the polypeptide be encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1-112.

Polypeptides of the present invention are also inclusive of derivative polypeptides encoded by a nucleic acid molecule according to the instant invention. In this embodiment, it is preferred that the polypeptide be a CSP. Also preferred are derivative polypeptides having an amino acid sequence selected from the group consisting of SEQ ID NO: 113-259 and which has been acetylated, carboxylated, phosphorylated, glycosylated, ubiquitinated or post-translationally modified in another manner. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H. In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications.

See, e.g., expasy.org (accessed November 11, 2002) of the world wide web, which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications include, but are not limited to: (Z)-dehydrobutyrine; 1-chondroitin sulfate-L-aspartic acid ester; 1'-glycosyl-L-tryptophan; 1'-phospho-L-histidine; 1-thioglycine; 2'-(S-L-cysteinyl)-L-histidine; 2'-[3-carboxamido (trimethylammonio)propyl]-L-histidine; 2'-alpha-mannosyl-L-tryptophan; 2-methyl-L-glutamine; 2-oxobutanoic acid; 2-pyrrolidone carboxylic acid; 3'-(1'-L-histidyl)-L-tyrosine; 3'-(8alpha-FAD)-L-histidine; 3'-(S-L-cysteinyl)-L-tyrosine; 3', 3",5'-triiodo-L-thyronine; 3'-4'-phospho-L-tyrosine; 3-hydroxy-L-proline; 3'-methyl-L-histidine; 3-methyl-L-lanthionine; 3'-phospho-L-histidine; 4'-(L-tryptophan)-L-tryptophyl quinone; 42 N-cysteinyl-glycosylphosphatidylinositolethanolamine; 43 -(T-L-histidyl)-L-tyrosine; 4-hydroxy-L-arginine; 4-hydroxy-L-lysine; 4-hydroxy-L-proline; 5'-(N6-L-lysine)-L-topaquinone; 5-hydroxy-L-lysine; 5-methyl-L-arginine; alpha-l-microglobulin-Ig alpha complex chromophore; bis-L-cysteinyl bis-L-histidino diiron disulfide; bis-L--cysteinyl-L-N3'-histidino-L-serinyl tetrairon' tetrasulfide; chondroitin sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; D-alanine; D-allo-isoleucine; D-asparagine; dehydroalanine; dehydrotyrosine; dermatan 4-sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; D-glucuronyl-N-glycine; dipyrrolylmethanemethyl-L-cysteine; D-leucine; D-methionine; D-phenylalanine; D-serine; D-tryptophan; glycine amide; glycine oxazolecarboxylic acid; glycine thiazolecarboxylic acid; heme P450-bis-L-cysteine-L-tyrosine; heme-bis-L-cysteine; hemediol-L-aspartyl ester-L-glutamyl ester; hemediol-L-aspartyl ester-L-glutamyl ester-L-methionine sulfonium; heme-L-cysteine; heme-L-histidine; heparan sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; heme P450-bis-L-cysteine-L-lysine; hexakis-L-cysteinyl hexairon hexasulfide; keratan sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-threonine; L oxoalanine- lactic acid; L phenyllactic acid; l'-(8alpha-FAD)-L-histidine; L-2'.4',5'-topaquinone; L-3',4'-dihydroxyphenylalanine; L-3'.4'.5'-trihydroxyphenylalanine; L-4'-

bromophenylalanine; L-6'-bromotryptophan; L-alanine amide; L-alanyl imidazolinone glycine; L-allysine; L-arginine amide; L-asparagine amide; L-aspartic 4-phosphoric anhydride; L-aspartic acid 1-amide; L-beta-methylthioaspartic acid; L-bromohistidine; L-citrulline; L-cysteine amide; L-cysteine glutathione disulfide; L-cysteine methyl disulfide; L-cysteine methyl ester; L-cysteine oxazolecarboxylic acid; L-cysteine oxazolinecarboxylic acid; L-cysteine persulfide; L-cysteine sulfenic acid; L-cysteine sulfinic acid; L-cysteine thiazolecarboxylic acid; L-cysteinyl homocitryl molybdenum-heptairon-nonasulfide; L-cysteinyl imidazolinone glycine; L-cysteinyl molybdopterin; L-cysteinyl molybdopterin guanine dinucleotide; L-cystine; L-erythro-beta-hydroxyasparagine; L-erythro-beta-hydroxyaspartic acid; L-gamma-carboxyglutamic acid; L-glutamic acid 1-amide; L-glutamic acid 5-methyl ester; L-glutamine amide; L-glutamyl 5-glycerylphosphorylethanolamine; L-histidine amide; L-isoglutamyl-polyglutamic acid; L-isoglutamyl-polyglycine; L-isoleucine amide; L-lanthionine; L-leucine amide; L-lysine amide; L-lysine thiazolecarboxylic acid; L-lysinoalanine; L-methionine amide; L-methionine sulfone; L-phenyalanine thiazolecarboxylic acid; L-phenylalanine amide; L-proline amide; L-selenocysteine; L-selenocysteinyl molybdopterin guanine dinucleotide; L-serine amide; L-serine thiazolecarboxylic acid; L-seryl imidazolinone glycine; L-T-bromophenylalanine; L-T-bromophenylalanine; L-threonine amide; L-thyroxine; L-tryptophan amide; L-tryptophyl quinone; L-tyrosine amide; L-valine amide; meso-lanthionine; N-(L-glutamyl)-L-tyrosine; N-(L-isoaspartyl)-glycine; N-(L-isoaspartyl)-L-cysteine; N,N,N-trimethyl-L-alanine; N,N-dimethyl-L-proline; N2-acetyl-L-lysine; N2-succinyl-L-tryptophan; N4-(ADP-ribosyl)-L-asparagine; N4-glycosyl-L-asparagine; N4-hydroxymethyl-L-asparagine; N4-methyl-L-asparagine; N5-methyl-L-glutamine; N6- 1 -carboxyethyl-L-lysine; N6-(4-amino hydroxybutyl)-L-lysine; N6-(L-isoglutamyl)-L-lysine; N6-(phospho-5'-adenosine)-L-lysine; N6-(phospho-5'-guanosine)-L-tysine; N6,N6,N6-trimethyl-L-lysine; N6,N6-dimethyl-L-lysine; N6-acetyl-L-lysine; N6-biotinyl-L-lysine; N6-carboxy-L-lysine; N6-formyl-L-lysine; N6-glycyl-L-lysine; N6-lipoyl-L-lysine; N6-methyl-L-lysine; N6-methyl-N6-poly(N-methyl-propylamine)-L-lysine; N6-mureinyl-L-lysine; N6-myristoyl-L-lysine; N6-palmitoyl-L-lysine; N6-pyridoxal phosphate-L-lysine; N6-pyruvic acid 2-iminyl-L-lysine; N6-retinal-L-lysine; N-acetylglycine; N-acetyl-L-glutamine; N-acetyl-L-alanine; N-acetyl-L-aspartic acid; N-acetyl-L-cysteine; N-acetyl-L-glutamic acid; N-acetyl-L-isoleucine; N-acetyl-L-methionine; N-acetyl-L-proline; N-acetyl-L-serine; N-acetyl-L-threonine; N-acetyl-L-

tyrosine; N-acetyl-L-valine; N-alanyl-glycosylphosphatidylinositolethanolamine; N-asparaginyl-glycosylphosphatidylinositolethanolamine; N-aspartyl-glycosylphosphatidylinositolethanolamine; N-formylglycine; N-formyl-L-methionine; N-glycyl-glycosylphosphatidylinositolethanolamine; N-L-glutamyl-poly-L-glutamic acid; N-methylglycine; N-methyl-L-alanine; N-methyl-L-methionine; N-methyl-L-phenylalanine; N-myristoyl-glycine; N-palmitoyl-L-cysteine; N-pyruvic acid 2-iminyl-L-cysteine; N-pyruvic acid 2-iminyl-L-valine; N-seryl-glycosylphosphatidylinositolethanolamine; N-seryl-glycosyCSPhingolipidinositolethanolamine; O-(ADP-ribosyl)-L-serine; O-(phospho-5'-adenosine)-L-threonine; O-(phospho-5'-DNA)-L-serine; O-(phospho-5'-DNA)-L-threonine; O-(phospho-5'rRNA)-L-serine; O-(phosphoribosyl dephospho-coenzyme A)-L-serine; O-(sn-l-glycerophosphoryl)-L-serine; O4'-(8alpha-FAD)-L-tyrosine; O4'-(phospho-5'-adenosine)-L-tyrosine; O4'-(phospho-5'-DNA)-L-tyrosine; O4'-(phospho-5'-RNA)-L-tyrosine; O4'-(phospho-5'-uridine)-L-tyrosine; O4-glycosyl-L-hydroxyproline; O4'-glycosyl-L-tyrosine; O4'-sulfo-L-tyrosine; O5-glycosyl-L-hydroxylysine; O-glycosyl-L-serine; O-glycosyl-L-threonine; omega-N-(ADP-ribosyl)-L-arginine; omega-N-omega-N'-dimethyl-L-arginine; omega-N-methyl-L-arginine; omega-N-omega-N-dimethyl-L-arginine; omega-N-phospho-L-arginine; O'octanoyl-L-serine; O-palmitoyl-L-serine; O-palmitoyl-L-threonine; O-phospho-L-serine; O-phospho-L-threonine; O-phosphopantetheine-L-serine; phycoerythrobilin-bis-L-cysteine; phycourobilin-bis-L-cysteine; pyrroloquinoline quinone; pyruvic acid; S hydroxycinnamyl-L-cysteine; S-(2-aminovinyl) methyl-D-eysteine; S-(2-aminovinyl)-D-cysteine; S-(6-FW-L-cysteine; S-(8alpha-FAD)-L-cysteine; S-(ADP-ribosyl)-L-cysteine; S-(L-isoglutamyl)-L-cysteine; S-12-hydroxyfarnesyl-L-cysteine; S-acetyl-L-cysteine; S-diacylglycerol-L-cysteine; S-diphytanylglycerot diether-L-cysteine; S-farnesyl-L-cysteine; S-geranylgeranyl-L-cysteine; S-glycosyl-L-cysteine; S-glycyl-L-cysteine; S-methyl-L-cysteine; S-nitrosyl-L-cysteine; S-palmitoyl-L-cysteine; S-phospho-L-cysteine; S-phycobiliviolin-L-cysteine; S-phycocyanobilin-L-cysteine; S-phycoerythrobilin-L-cysteine; S-phytochromobilin-L-cysteine; S-selenyl-L-cysteine; S-sulfo-L-cysteine; tetrakis-L-cysteinyl diiron disulfide; tetrakis-L-cysteinyl iron; tetrakis-L-cysteinyl tetrairon tetrasulfide; trans-2,3-cis 4-dihydroxy-L-proline; tris-L-cysteinyl triiron tetrasulfide; tris-L-cysteinyl triiron trisulfide; tris-L-cysteinyl-L-aspartato tetrairon tetrasulfide; tris-L-cysteinyl-L-cysteine persulfido-bis-L-glutamato-L-histidino tetrairon disulfide trioxide; tris-L-cysteinyl-L-N3'-histidino

tetrairon tetrasulfide; tris-L-cysteinyl-L-Nl'-histidino tetrairon tetrasulfide; and tris-L-cysteinyl-L-serinyl tetrairon tetrasulfide.

Additional examples of PTMs may be found in web sites such as the Delta Mass database based on Krishna, R. G. and F. Wold (1998). Posttranslational Modifications. Proteins - Analysis and Design. R. H. Angeletti. San Diego, Academic Press. 1: 121-206; Methods in Enzymology, 193, J.A. McClosky (ed) (1990), pages 647-660; Methods in Protein Sequence Analysis edited by Kazutomo Imahori and Fumio Sakiyama, Plenum Press, (1993) "Post-translational modifications of proteins" R.G. Krishna and F. Wold pages 167-172; "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. Nucleic Acids Res. 29; 332-335 (2001) "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. Nucleic Acids Research, 27: 370-372 (1999); and "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al.Nucleic Acids Res 27(1):237-239 (1999) see also,WO 02/21139A2, the disclosure of which is incorporated herein by reference in its entirety.

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signalling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical

analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g, p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website expasy.org of the world wide web. The nucleic acid molecule may also be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing events and events brought about by human manipulation which do not occur naturally. Circular, branched and branched

circular polypeptides may be synthesized by non-translation natural processes and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB,

DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

Polypeptides of the present invention, including full length polypeptides, fragments and fusion proteins, can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to polypeptides of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

Polypeptides of the present invention, including full length polypeptides, fragments and fusion proteins, can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-CSP antibodies.

Polypeptides of the present invention, including full length polypeptides, fragments and fusion proteins, can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999). PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

Polypeptides of the present invention are also inclusive of analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, this polypeptide is a CSP. In a more preferred embodiment, this polypeptide is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 113-259. Also preferred is an analog polypeptide comprising one or more

substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally occurring polypeptide. In one embodiment, the analog is structurally similar to a CSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --CH(OH)CH_2-- and --CH_2SO--. In another embodiment, the analog comprises substitution of one or more amino acids of a CSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g.*, Kole *et al.*, *Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.

Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, *inter alia*, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (1993).

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinoyl-(9-fluorenylmethoxycarbonyl)-L-lysine (FMOC biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of an *E. coli* BirA substrate peptide. The FMOC and *t*BOC derivatives of dabcyl-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyl chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyl quencher in fluorescence resonance energy transfer (FRET) systems, can be

introduced during automated synthesis of peptides by using EDANS-FMOC-L-glutamic acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated FMOC synthesis of peptides using (FMOC)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other FMOC-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)-β-alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-

pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

Fusion Proteins

Another aspect of the present invention relates to the fusion of a polypeptide of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide of the present invention is a CSP. In a more preferred embodiment, the polypeptide of the present invention that is fused to a heterologous polypeptide which comprises part or all of the amino acid sequence of SEQ ID NO: 113-259, or is a mutein, homologous polypeptide, analog or derivative thereof. In an even more preferred embodiment, the fusion protein is encoded by a nucleic acid molecule comprising all or part of the nucleic acid sequence of SEQ ID NO: 1-112, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-112.

The fusion proteins of the present invention will include at least one fragment of a polypeptide of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the polypeptide of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of a polypeptide of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and preferably at least 15, 20, or 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particularly useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g.*, Ausubel, Chapter 16, (1992), *supra*. Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins into the periplasmic space or extracellular milieu for prokaryotic hosts or into the culture medium for eukaryotic cells through incorporation of secretion signals and/or leader sequences. For example, a His^6 tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful fusion proteins of the present invention include those that permit use of the polypeptide of the present invention as bait in a yeast two-hybrid system. *See* Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.*

10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); Colas *et al.*, *Nature* 380, 548-550 (1996); Norman, T. *et al.*, *Science* 285, 591-595 (1999); Fabbrizio *et al.*, *Oncogene* 18, 4357-4363 (1999); Xu *et al.*, *Proc Natl Acad Sci U S A.* 94, 12473-12478 (1997); Yang, *et al.*, *Nuc. Acids Res.* 23, 1152-1156 (1995); Kolonin *et al.*, *Proc Natl Acad Sci U S A* 95, 14266-14271 (1998); Cohen *et al.*, *Proc Natl Acad Sci U S A* 95, 14272-14277 (1998); Uetz, *et al. Nature* 403, 623-627(2000); Ito, *et al.*, *Proc Natl Acad Sci U S A* 98, 4569-4574 (2001). Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded polypeptide on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above.

The polypeptides of the present invention can also usefully be fused to protein toxins, such as Pseudomonas exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, or ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins, β-galactosidase, biotin trpE, protein A, β-lactamase, α-amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast α mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. *See*, *e.g.*, Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well known in the art (*e.g.*, a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the CSP.

As further described below, the polypeptides of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize polypeptides of the present invention including CSPs and their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly CSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of CSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of CSPs.

One may determine whether polypeptides of the present invention including CSPs, muteins, homologous proteins or allelic variants or fusion proteins of the present invention are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the polypeptide at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); and combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, (Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides or fusion proteins of the present invention is well known and within the skill of one having ordinary skill in the art. *See, e.g.*, Scopes, <u>Protein Purification</u>, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated polypeptides or fusion proteins of the present invention in pure or substantially pure form

in the presence or absence of a stabilizing agent. Stabilizing agents include both proteinaceous and non-proteinaceous material and are well known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

Although high levels of purity are preferred when the isolated polypeptide or fusion protein of the present invention are used as therapeutic agents, such as in vaccines and replacement therapy, the isolated polypeptides of the present invention are also useful at lower purity. For example, partially purified polypeptides of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In a preferred embodiment, the purified and substantially purified polypeptides of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides or fusion proteins of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent. For example, the peptides of the invention may be stabilized by covalent linkage to albumin. See, U.S. Patent No. 5,876,969, the contents of which are hereby incorporated in its entirety.

The polypeptides or fusion proteins of the present invention can also be usefully bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the polypeptides or fusion proteins of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized polypeptide or fusion protein of the present invention.

As another example, the polypeptides or fusion proteins of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides and fusion proteins of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so

attached, the polypeptide or fusion protein of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound polypeptide or fusion protein to indicate biologic interaction there between. The polypeptides or fusion proteins of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the polypeptide or fusion protein of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound polypeptide or fusion protein to indicate biological interaction there between.

Alternative Transcripts

In antother aspect, the present invention provides splice variants of genes and proteins encoded thereby. The identification of a novel splice variant which encodes an amino acid sequence with a novel region can be targeted for the generation of reagents for use in detection and/or treatment of cancer. The novel amino acid sequence may lead to a unique protein structure, protein subcellular localization, biochemical processing or function of the splice variant. This information can be used to directly or indirectly facilitate the generation of additional or novel therapeutics or diagnostics. The nucleotide sequence in this novel splice variant can be used as a nucleic acid probe for the diagnosis and/or treatment of cancer.

Specifically, the newly identified sequences may enable the production of new antibodies or compounds directed against the novel region for use as a therapeutic or diagnostic. Alternatively, the newly identified sequences may alter the biochemical or biological properties of the encoded protein in such a way as to enable the generation of improved or different therapeutics targeting this protein.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention. In a preferred embodiment, the antibodies are specific for a polypeptide that is a CSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 113-259, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may also be due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a CSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or vice versa. In addition, alternative splice forms of a CSP may be indicative of cancer. Differential degradation of the C or N-terminus of a CSP may also be a marker or target for anticancer therapy. For example, a CSP may be N-terminal degraded in cancer cells exposing new epitopes to antibodies which may selectively bind for diagnostic or therapeutic uses.

As is well known in the art, the degree to which an antibody can discriminate among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-CSP polypeptides by at least two-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the polypeptide of the present invention in samples derived from human colon.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to $1 \text{ X } 10^{-13}$ M proving especially useful.

The antibodies of the present invention can be naturally occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In such case, antibodies to the polypeptides of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the polypeptide of the present invention. Such antibodies will typically, but will not

invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patent Nos. 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention are also usefully obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster), lagomorphs (typically rabbits), and also larger mammals, such as sheep, goats, cows, and horses; or egg laying birds or reptiles such as chickens or alligators. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the polypeptide of the present invention. One form of avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, which is herein incorporated by reference in its entirety.

As discussed above, virtually all fragments of 8 or more contiguous amino acids of a polypeptide of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptides of the present invention to other moieties. For example, polypeptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. *See* Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996). Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization. Moss, *Semin. Immunol.* 2: 317-327 (1990).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the polypeptides of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the polypeptides of the present invention. Antibodies from avian species may have particular advantage in detection of the polypeptides of the present invention, in human serum or tissues. Vikinge et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998). Following immunization, the antibodies of the present invention can be obtained using any art-accepted technique. Such techniques are well known in the art and are described in detail in references such as Coligan, *supra*; Zola, *supra*; Howard *et al.* (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); and Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997).

Briefly, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two

methods of production are not mutually exclusive: genes encoding antibodies specific for the polypeptides of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the polypeptides of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S. Patent No. 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant antibody production of whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.*, Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*, 4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et al.*, *Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994). Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.*, Barbas (2001), *supra*; Kay, *supra*; and Abelson, *supra*.

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell. Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention. For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.*, Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, J. Biotechnol. 76(2-3):1 57-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20

(1999); Pennell *et al., Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al., J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al., Res. Immunol.* 149(6): 589-99 (1998); and Shusta *et al., Nature Biotechnol.* 16(8): 773-7 (1998).

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.,* Li *et al., Protein Expr. Purif.* 21(1): 121-8 (2001); Ailor *et al., Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al., Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al., Immunology* 91(1): 13-9 (1997); and Nesbit *et al., J. Immunol. Methods* 151(1-2): 201-8 (1992).

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al., Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al., Biotechniques* 29(1): 128-38 (2000); Fischer *et al., J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al., Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al., Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al., Plant Physiol.* 109(2): 341-6 (1995).

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock et al., *J. Immunol Methods.* 231: 147-57 (1999); Young et al., *Res. Immunol.* 149: 609-10 (1998); and Limonta et al., *Immunotechnology* 1: 107-13 (1995).

Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells. Verma *et al., J. Immunol. Methods* 216(1-2):165-81 (1998) review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies. Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al., J. Biochem.* (Tokyo) 125(2): 328-33 (1999) and Ryabova *et al., Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al., J. Immunol. Methods* 231(1-2): 147-57 (1999).

The invention further provides antibody fragments that bind specifically to one or more of the polypeptides of the present invention or to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid

molecules of the present invention. Among such useful fragments are Fab, Fab', Fv, $F(ab)'_2$, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

The present invention also relates to antibody derivatives that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus are more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful method is PEGylation to increase the serum half life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, Morrison *et al.*, *Proc. Natl. Acad. Sci USA*.81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985); and U.S. Patent No. 5,807,715 the disclosure of which is incorporated herein by reference in its entirety. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); and U.S. Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties. Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. Accordingly, the present invention includes any recombinant vector containing the coding sequences, or part

thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., Proc. Natl. Acad. Sci. (USA) 90: 7889-7893 (1993); Duan et al., Proc. Natl. Acad. Sci. (USA) 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention. The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label can usefully be an enzyme that catalyzes production and local deposition of a detectable product. Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well known, and include alkaline phosphatase, β-galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopryanoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish

peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. *See, e.g.*, Thorpe *et al.*, *Methods Enzymol.* 133: 331-53 (1986); Kricka *et al.*, *J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al.*, *J. Biolumin. Chemilumin.* 10(6): 353-9 (1995). Kits for such enhanced chemiluminescent detection (ECL) are available commercially. The antibodies can also be labeled using colloidal gold.

As another example, when the antibodies of the present invention are used, *e.g.*, for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores. There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention. For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention. For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P, ^{32}P, ^{35}S, ^{3}H, and ^{125}I. As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th, ^{227}Ac, ^{225}Ac, ^{223}Ra, ^{213}Bi, ^{212}Pb, ^{212}Bi, ^{211}At, ^{203}Pb, ^{194}Os, ^{188}Re, ^{186}Re, ^{153}Sm, ^{149}Tb, ^{131}I, ^{125}I, ^{111}In, ^{105}Rh, ^{99m}Tc, ^{97}Ru, ^{90}Y, ^{90}Sr, ^{88}Y, ^{72}Se, ^{67}Cu, or ^{47}Sc.

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application as for which they were mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the polypeptides of the present invention. Commonly, the antibody in such immunotoxins is conjugated to Pseudomonas exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, or ricin. *See* Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998).

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, attached to a substrate. Substrates can be porous or nonporous, planar or nonplanar. For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography. For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microsphere can then be used for isolation of cells that express or display the polypeptides of the present invention. As

another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the CSPs of the present invention or to polypeptides encoded by the CSNAs of the invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding a CSP. In a preferred embodiment, the CSP comprises an amino acid sequence selected from SEQ ID NO: 113-259, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a CSNA of the invention, preferably a CSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-112, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human CSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well known in the art. *See*, *e.g.*, Hogan *et*

al., Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g.*, Paterson *et al.*, *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and U.S. Patent No. 4,873,191, herein incorporated by reference in its entirety); retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g.*, Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g.*, Thompson *et al.*, *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g.*, Lo, 1983, *Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g.*, Ulmer *et al.*, *Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g.*, Lavitrano *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.*, Campell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e.*, a nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, i.e. e., mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, e.g., the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232- 6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the

transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g.*, Gu *et al.*, *Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g.*, Smithies *et al.*, *Nature* 317: 230-234 (1985); Thomas *et al.*, *Cell* 51: 503-512 (1987); Thompson *et al.*, *Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention in vivo. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g.,* Thomas, *supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site in vivo using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.*, knockouts) are administered to a patient in vivo. Such cells may be obtained from an animal or patient or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered in vitro using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.*, U.S. Patent Nos. 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention is a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 113-259 and SEQ ID NO: 1-112 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation,

chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence. In addition, the invention includes a method of using patterns of expression associated with either the nucleic acids or proteins in a computer-based method to diagnose disease.

Diagnostic Methods for Colon Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by comparing expression of a CSNA or a CSP in a human patient that has or may have colon cancer, or who is at risk of developing colon cancer, with the expression of a CSNA or a CSP in a normal human control. For purposes of the present invention, "expression of a CSNA" or "CSNA expression" means the quantity of CSNA mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of a CSP" or "CSP expression" means the amount of CSP that can be measured by any method known in the art or the level of translation of a CSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing colon cancer in a patient, in particular adenocarcinoma, by analyzing for changes in levels of CSNA or CSP in cells, tissues, organs or bodily fluids compared with levels of CSNA or CSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of a CSNA or CSP in the patient versus the normal human control is associated with the presence of colon cancer or with a predilection to the disease. In another preferred embodiment, the present invention

provides methods for diagnosing colon cancer in a patient by analyzing changes in the structure of the mRNA of a CSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing colon cancer in a patient by analyzing changes in a CSP compared to a CSP from a normal patient. These changes include, *e.g.*, alterations, including post translational modifications such as glycosylation and/or phosphorylation of the CSP or changes in the subcellular CSP localization.

For purposes of the present invention, diagnosing means that CSNA or CSP levels are used to determine the presence or absence of disease in a patient. As will be understood by those of skill in the art, measurement of other diagnostic parameters may be required for definitive diagnosis or determination of the appropriate treatment for the disease. The determination may be made by a clinician, a doctor, a testing laboratory, or a patient using an over the counter test. The patient may have symptoms of disease or may be asymptomatic. In addition, the CSNA or CSP levels of the present invention may be used as screening marker to determine whether further tests or biopsies are warranted. In addition, the CSNA or CSP levels may be used to determine the vulnerability or susceptibility to disease.

In a preferred embodiment, the expression of a CSNA is measured by determining the amount of a mRNA that encodes an amino acid sequence selected from SEQ ID NO: 113-259, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the CSNA expression that is measured is the level of expression of a CSNA mRNA selected from SEQ ID NO: 1-112, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acid molecules. CSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See*, *e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. CSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of a CSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*, aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, CSNA expression

may be compared to a known control, such as normal colon nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of a CSP is measured by determining the level of a CSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 113-259, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of a CSNA or CSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of colon cancer. The expression level of a CSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the CSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See*, e.g, Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the CSP structure may be determined by any method known in the art, including, *e.g.*, using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to a CSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-CSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the CSP will bind to the anti-CSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-CSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the CSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of a CSP in the

sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure CSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-CSP antibody is attached to a solid support and an allocated amount of a labeled CSP and a sample of interest are incubated with the solid support. The amount of labeled CSP attached to the solid support can be correlated to the quantity of a CSP in the sample.

Of the proteomic approaches, 2D PAGE is a well known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of a CSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more CSNAs of interest. In this approach, all or a portion of one or more CSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur

between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. As used herein "blood" includes whole blood, plasma, serum, circulating epithelial cells, constituents, or any derivative of blood.

In addition to detection in bodily fluids, the proteins and nucleic acids of the invention are suitable to detection by cell capture technology. Whole cells may be captured by a variety methods for example magnetic separation, such as described in U.S. Patent. Nos. 5,200,084; 5,186,827; 5,108,933; and 4,925,788, the disclosures of which are incorporated herein by reference in their entireties. Epithelial cells may be captured using such products as Dynabeads® or CELLection™ (Dynal Biotech, Oslo, Norway). Alternatively, fractions of blood may be captured, e.g., the buffy coat fraction (50mm cells isolated from 5ml of blood) containing epithelial cells. In addition, cancer cells may be captured using the techniques described in WO 00/47998, the disclosure of which is incorporated herein by reference in its entirety. Once the cells are captured or concentrated, the proteins or nucleic acids are detected by the means described in the subject application. Alternatively, nucleic acids may be captured directly from blood samples, see U.S. Patent Nos. 6,156,504, 5,501,963; or WO 01/42504 , the disclosures of which are incorporated herein by reference in their entireties.

In a preferred embodiment, the specimen tested for expression of CSNA or CSP includes without limitation colon tissue, fecal samples, colonocytes, colon cells grown in cell culture, blood, serum, lymph node tissue, and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary colon cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, lungs, and adrenal glands. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediatinoscopy,

endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration.

Colonocytes represent an important source of the CSP or CSNA because they provide a picture of the immediate past metabolic history of the GI tract of a subject. In addition, such cells are representative of the cell population from a statistically large sampling frame reflecting the state of the colonic mucosa along the entire length of the colon in a non-invasive manner, in contrast to a limited sampling by colonic biopsy using an invasive procedure involving endoscopy. Specific examples of patents describing the isolation of colonocytes include U.S. Patent Nos. 6,335,193; 6,020,137 5,741,650; 6,258,541; US 2001 0026925 A1; WO 00/63358 A1, the disclosures of which are incorporated herein by reference in their entireties.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of a CSNAs or CSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other CSNA or CSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular CSNA or CSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more CSNA and/or CSP in a sample from a patient suspected of having colon cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of a CSNA and/or CSP and then ascertaining whether the patient has colon cancer from the expression level of the CSNA or CSP. In general, if high expression relative to a control of a CSNA or CSP is indicative of colon cancer, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least one and a half times higher, and more preferably are at least two times higher, still more preferably five

times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is indicative of colon cancer, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least one and a half times lower, and more preferably are at least two times lower, still more preferably five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether colon cancer has metastasized in a patient. One may identify whether the colon cancer has metastasized by measuring the expression levels and/or structural alterations of one or more CSNAs and/or CSPs in a variety of tissues. The presence of a CSNA or CSP in a tissue other than colon at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of a CSNA or CSP is associated with colon cancer. Similarly, the presence of a CSNA or CSP in a tissue other than colon at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of a CSNA or CSP is associated with colon cancer. Further, the presence of a structurally altered CSNA or CSP that is associated with colon cancer is also indicative of metastasis.

In general, if high expression relative to a control of a CSNA or CSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the CSNA or CSP is at least one and a half times higher, and more preferably are at least two times higher, still more preferably five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the CSNA or CSP is at least one and a half times lower, and more preferably are at least two times lower, still more preferably five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

Staging

The invention also provides a method of staging colon cancer in a human patient. The method comprises identifying a human patient having colon cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more CSNAs or CSPs. First, one or more tumors from a variety of patients are staged according to procedures well known in the art, and the expression levels of one or more CSNAs or CSPs is determined for each stage to obtain a standard expression level for each CSNA and CSP. Then, the CSNA or CSP expression levels of the CSNA or CSP are determined in a biological sample from a patient whose stage of cancer is not known. The CSNA or CSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the CSNAs and CSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of a CSNA or CSP to determine the stage of a colon cancer.

Monitoring

Further provided is a method of monitoring colon cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the colon cancer. The monitoring may determine if there has been a reoccurrence and, if so, determine its nature. The method comprises identifying a human patient that one wants to monitor for colon cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more CSNAs or CSPs, and comparing the CSNA or CSP levels over time to those CSNA or CSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a CSNA or CSP that are associated with colon cancer.

If increased expression of a CSNA or CSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of a CSNA or CSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased

expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of a CSNA or CSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting a decrease in the expression level of a CSNA or CSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of CSNAs or CSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples. Monitoring a patient for onset of colon cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a CSNA and/or CSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more CSNAs and/or CSPs are detected. The presence of higher (or lower) CSNA or CSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly colon cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more CSNAs and/or CSPs of the invention can also be monitored by analyzing levels of expression of the CSNAs and/or CSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

The methods of the present invention can also be used to detect genetic lesions or mutations in a CSG, thereby determining if a human with the genetic lesion is susceptible to developing colon cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing colon cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the CSGs of this invention, a chromosomal rearrangement of a CSG, an aberrant modification of a CSG (such as of the methylation pattern of the genomic DNA), or allelic loss of a CSG. Methods to detect such lesions in the CSG of

this invention are known to those having ordinary skill in the art following the teachings of the specification.

Methods of Detecting Noncancerous Colon Diseases

The present invention also provides methods for determining the expression levels and/or structural alterations of one or more CSNAs and/or CSPs in a sample from a patient suspected of having or known to have a noncancerous colon disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of a CSNA and/or CSP, comparing the expression level or structural alteration of the CSNA or CSP to a normal colon control, and then ascertaining whether the patient has a noncancerous colon disease. In general, if high expression relative to a control of a CSNA or CSP is indicative of a particular noncancerous colon disease, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is indicative of a noncancerous colon disease, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether a CSNA and/or CSP is associated with a particular noncancerous colon disease by obtaining colon tissue from a patient having a noncancerous colon disease of interest and determining which CSNAs and/or CSPs are expressed in the tissue at either a higher or a lower level than in normal colon tissue. In another embodiment, one may determine whether a CSNA or CSP exhibits structural alterations in a particular noncancerous colon disease state by obtaining colon tissue from a patient having a noncancerous colon disease of interest and determining the structural alterations in one or more CSNAs and/or CSPs relative to normal colon tissue.

Methods for Identifying Colon Tissue

In another aspect, the invention provides methods for identifying colon tissue. These methods are particularly useful in, *e.g.*, forensic science, colon cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is colon tissue or has colon tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising colon tissue or having colon tissue-like characteristics, determining whether the sample expresses one or more CSNAs and/or CSPs, and, if the sample expresses one or more CSNAs and/or CSPs, concluding that the sample comprises colon tissue. In a preferred embodiment, the CSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 113-259, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the CSNA has a nucleotide sequence selected from SEQ ID NO: 1-112, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses a CSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether a CSP is expressed. Determining whether a sample expresses a CSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the CSP has an amino acid sequence selected from SEQ ID NO: 113-259, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two CSNAs and/or CSPs is determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five CSNAs and/or CSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is colon tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into colon tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new colon tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation

include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

Methods for Producing and Modifying Colon Tissue

In another aspect, the invention provides methods for producing engineered colon tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing a CSNA or a CSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of colon tissue cells. In a preferred embodiment, the cells are pleuripotent. As is well known in the art, normal colon tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered colon tissue or cells comprises one of these cell types. In another embodiment, the engineered colon tissue or cells comprises more than one colon cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the colon cell tissue. Methods for manipulating culture conditions are well known in the art.

Nucleic acid molecules encoding one or more CSPs are introduced into cells, preferably pleuripotent cells. In a preferred embodiment, the nucleic acid molecules encode CSPs having amino acid sequences selected from SEQ ID NO: 113-259, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1-112, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, a CSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well known in the art and are described in detail, *supra*.

Artificial colon tissue may be used to treat patients who have lost some or all of their colon function.

Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, fusion proteins, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, or inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises a CSNA or part thereof. In a more preferred embodiment, the CSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-112, a nucleic acid that hybridizes thereto, an allelic

variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises a CSP or fragment thereof. In a more preferred embodiment, the pharmaceutical composition comprises a CSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 113-259, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-CSP antibody, preferably an antibody that specifically binds to a CSP having an amino acid that is selected from the group consisting of SEQ ID NO: 113-259, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Due to the association of angiogenesis with cancer vascularization there is great need of new markers and methods for diagnosing angiogenesis activity to identify developing tumors and angiogenesis related diseases. Furthermore, great need is also present for new molecular targets useful in the treatment of angiogenesis and angiogenesis related diseases such as cancer. In addition known modulators of angiogenesis such as endostatin or vascular endothelial growth factor (VEGF). Use of the methods and compositions disclosed herein in combination with anti-angiogenesis drugs, drugs that block the matrix breakdown (such as BMS-275291, Dalteparin (Fragmin®), Suramin), drugs that inhibit endothelial cells (2-methoxyestradiol (2-ME), CC-5013 (Thalidomide Analog), Combretastatin A4 Phosphate, LY317615 (Protein Kinase C Beta Inhibitor), Soy Isoflavone (Genistein; Soy Protein Isolate), Thalidomide), drugs that block activators of angiogenesis (AE-941 (Neovastat™; GW786034), Anti-VEGF Antibody (Bevacizumab; Avastin™), Interferon-alpha, PTK787/ZK 222584, VEGF-Trap, ZD6474), Drugs that inhibit endothelial-specific integrin/survival signaling (EMD 121974, Anti-Anb3 Integrin Antibody (Medi-522; Vitaxin™)).

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art that is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.),

Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000) and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, cornstarch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate,

isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically. For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example CSP polypeptide, fusion protein, or fragments thereof, antibodies specific for CSP, agonists, antagonists or inhibitors of CSP, which ameliorates the signs or symptoms of the disease or prevent progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

<u>Therapeutic Methods</u>

The present invention further provides methods of treating subjects having defects in a gene of the invention, *e.g.*, in expression, activity, distribution, localization, and/or

solubility, which can manifest as a disorder of colon function. As used herein, "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for the purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further described in U.S. Patent Nos. 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g.*, Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid molecule of the present invention is administered. The nucleic acid molecule can be delivered in a vector that drives expression of a CSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a CSP are administered, for example, to complement a deficiency in the native CSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g.*, Cid-Arregui, *supra*. In a preferred embodiment, the nucleic acid molecule encodes a CSP having the amino acid sequence of SEQ ID NO: 113-259, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a CSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in CSP

production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode a CSP having the amino acid sequence of SEQ ID NO: 113-259, or a fragment, fusion protein, allelic variant or homolog thereof.

Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of a CSG antisense nucleic acid, are administered to downregulate transcription and/or translation of a CSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of a CSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to CSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995).

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the CSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); and McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000). Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding a CSP, preferably a CSP comprising an amino acid sequence of SEQ ID NO: 113-259, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-112, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a CSP, a fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant CSP defect.

Protein compositions are administered, for example, to complement a deficiency in native CSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to CSP. The immune response can be used to modulate activity of CSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate CSP.

In a preferred embodiment, the polypeptide administered is a CSP comprising an amino acid sequence of SEQ ID NO: 113-259, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-112, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well known, antibody compositions are administered, for example, to antagonize activity of CSP, or to target therapeutic agents to sites of CSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to a CSP comprising an amino acid sequence of SEQ ID NO: 113-259, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to a CSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-112, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to a CSP or have a modulatory effect on the expression or activity of a CSP. Modulators which decrease the expression or activity of CSP (antagonists) are believed to be useful in treating colon cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules

predicted via computer imaging to specifically bind to regions of a CSP can also be designed, synthesized and tested for use in the imaging and treatment of colon cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the CSPs identified herein. Molecules identified in the library as being capable of binding to a CSP are key candidates for further evaluation for use in the treatment of colon cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of a CSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of CSP is administered. Antagonists of CSP can be produced using methods generally known in the art. In particular, purified CSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of a CSP.

In other embodiments a pharmaceutical composition comprising an agonist of a CSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a CSP comprising an amino acid sequence of SEQ ID NO: 113-259, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a CSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-112, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting Colon Tissue

The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the colon or to specific cells in the colon. In a preferred embodiment, an anti-CSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if colon tissue needs to be selectively destroyed. This would be useful for targeting and killing colon cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting colon cell function.

In another embodiment, an anti-CSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring colon function, identifying colon cancer tumors, and identifying noncancerous colon diseases.

EXAMPLES

Example 1a: Alternative Splice Variants

We identified gene transcripts using the Gencarta™ tools (Compugen Ltd., Tel Aviv, Israel) and a variety of public and proprietary databases. These splice variants are either sequences which differ from a previously defined sequence or new uses of known sequences. In general related variants are annotated as DEX0449_XXX.nt.1, DEX0449_XXX.nt.2, DEX0449_XXX.nt.3, etc. The variant DNA sequences encode proteins which differ from a previously defined protein sequence. In relation to the nucleotide sequence naming convention, protein variants are annotated as DEX0449_XXX.aa.1, DEX0449_XXX.aa.2, etc., wherein transcript DEX0449_XXX.nt.1 encodes protein DEX0449_XXX.aa.1. A single transcript may encode a protein from an alternate Open Reading Fram (ORF) which is designated DEX0449_XXX.orf.1. Additionally, multiple transcripts may encode for a single protein. In this case, DEX0449_XXX.nt.1 and DEX0449_XXX.nt.2 will both be associated with DEX0449_XXX.aa.1.

The mapping of the nucleic acid ("NT") SEQ ID NO; DEX ID; chromosomal location (if known); open reading frame (ORF) location; amino acid ("AA") SEQ ID NO; AA DEX ID; are shown in the table below.

SEQ ID NO	DEX ID	Chromo Map	ORF Loc	SEQ ID NO	DEX ID
1	DEX0449_001.nt.1	17q25.3	1-498	113	DEX0449_001.aa.1
1	DEX0449_001.nt.1	17q25.3	2-493	114	DEX0449_001.orf.1
2	DEX0449_002.nt.1	5q13.1	253-1022	115	DEX0449_002.aa.1
2	DEX0449_002.nt.1	5q13.1	128-877	116	DEX0449_002.orf.1
3	DEX0449_003.nt.1	17q25.1	9-1226	117	DEX0449_003.aa.1
4	DEX0449_003.nt.2	17q25.1	7-1229	117	DEX0449_003.aa.1
5	DEX0449_003.nt.3	17q25.1	9-1583	118	DEX0449_003.aa.3
6	DEX0449_003.nt.4	17q25.1	2373-3357	119	DEX0449_003.aa.4
6	DEX0449_003.nt.4	17q25.1	2207-3358	120	DEX0449_003.orf.4

7	DEX0449_003.nt.5	17q25.1	7-1229	117	DEX0449_003.aa.1
8	DEX0449_003.nt.6	17q25.1	7-1229	117	DEX0449_003.aa.1
8	DEX0449_003.nt.6	17q25.1	2490-3839	121	DEX0449_003.orf.6
9	DEX0449_003.nt.7	17q25.1	9-914	122	DEX0449_003.aa.7
10	DEX0449_004.nt.1	4q32.3	237-722	123	DEX0449_004.aa.1
11	DEX0449_005.nt.1	8q24.3	181-802	124	DEX0449_005.aa.1
11	DEX0449_005.nt.1	8q24.3	3-1142	125	DEX0449_005.orf.1
12	DEX0449_005.nt.2	8q24.3	181-802	124	DEX0449_005.aa.1
12	DEX0449_005.nt.2	8q24.3	3-1142	126	DEX0449_005.orf.2
13	DEX0449_005.nt.3	8q24.3	229-1235	127	DEX0449_005.aa.3
13	DEX0449_005.nt.3	8q24.3	594-1574	128	DEX0449_005.orf.3
14	DEX0449_005.nt.4	8q24.3	180-662	129	DEX0449_005.aa.4
15	DEX0449_005.nt.5	8q24.3	181-667	129	DEX0449_005.aa.4
16	DEX0449_005.nt.6	8q24.3	181-802	124	DEX0449_005.aa.1
16	DEX0449_005.nt.6	8q24.3	3-980	130	DEX0449_005.orf.6
17	DEX0449_006.nt.1	14q32.33	1774-2421	131	DEX0449_006.aa.1
18	DEX0449_007.nt.1	4q13.3	103-431	132	DEX0449_007.aa.1
18	DEX0449_007.nt.1	4q13.3	2-427	133	DEX0449_007.orf.1
19	DEX0449_008.nt.1	4q21.1	232-1681	134	DEX0449_008.aa.1
19	DEX0449_008.nt.1	4q21.1	504-1679	135	DEX0449_008.orf.1
20	DEX0449_009.nt.1	4q21.1	1-951	136	DEX0449_009.aa.1
20	DEX0449_009.nt.1	4q21.1	3-944	137	DEX0449_009.orf.1
21	DEX0449_009.nt.2	4q21.1	48-554	138	DEX0449_009.aa.2
22	DEX0449_010.nt.1	15q25.3	117-561	139	DEX0449_010.aa.1
22	DEX0449_010.nt.1	15q25.3	61-558	140	DEX0449_010.orf.1
23	DEX0449_011.nt.1	8q11.23	868-1633	141	DEX0449_011.aa.1
23	DEX0449_011.nt.1	8q11.23	723-1628	142	DEX0449_011.orf.1
24	DEX0449_012.nt.1	7p11.2	733-1407	143	DEX0449_012.aa.1
25	DEX0449_012.nt.2	7p11.2	918-1596	143	DEX0449_012.aa.1
26	DEX0449_012.nt.3	7p11.2	733-1411	143	DEX0449_012.aa.1
27	DEX0449_013.nt.1	8p21.2	66-812	144	DEX0449_013.aa.1
28	DEX0449_014.nt.1	18p11.31	880-1218	145	DEX0449_014.aa.1
29	DEX0449_015.nt.1	20q13.2	422-727	146	DEX0449_015.aa.1
30	DEX0449_016.nt.1	7p22.3	394-1633	147	DEX0449_016.aa.1
30	DEX0449_016.nt.1	7p22.3	1019-1627	148	DEX0449_016.orf.1
31	DEX0449_017.nt.1	17p13.3	295-642	149	DEX0449_017.aa.1
31	DEX0449_017.nt.1	17p13.3	138-560	150	DEX0449_017.orf.1
32	DEX0449_018.nt.1	2q35	133-1414	151	DEX0449_018.aa.1

33	DEX0449_018.nt.2	2q35	133-1414	151	DEX0449_018.aa.1
34	DEX0449_018.nt.3	2q35	133-1414	151	DEX0449_018.aa.1
35	DEX0449_018.nt.4	2q35	134-1243	152	DEX0449_018.aa.4
36	DEX0449_018.nt.5	2q35	134-757	153	DEX0449_018.aa.5
37	DEX0449_018.nt.6	2q35	133-1414	151	DEX0449_018.aa.1
38	DEX0449_018.nt.7	2q35	2635-3261	154	DEX0449_018.orf.7
38	DEX0449_018.nt.7	2q35	133-511	155	DEX0449_018.aa.7
39	DEX0449_018.nt.8	2q35	134-1411	151	DEX0449_018.aa.1
40	DEX0449_019.nt.1	9	2070-2642	156	DEX0449_019.orf.1
40	DEX0449_019.nt.1	9	1311-1704	157	DEX0449_019.aa.1
41	DEX0449_020.nt.1	19q13.41	266-1285	158	DEX0449_020.orf.1
41	DEX0449_020.nt.1	19q13.41	364-1286	159	DEX0449_020.aa.1
42	DEX0449_020.nt.2	19q13.41	3-1457	160	DEX0449_020.orf.2
42	DEX0449_020.nt.2	19q13.41	280-1679	161	DEX0449_020.aa.2
43	DEX0449_021.nt.1	1q32.1	183-404	162	DEX0449_021.orf.1
43	DEX0449_021.nt.1	1q32.1	1-193	163	DEX0449_021.aa.1
44	DEX0449_022.nt.1	17q25.3	612-1550	164	DEX0449_022.aa.1
45	DEX0449_022.nt.2	17q25.3	72-1256	165	DEX0449_022.orf.2
45	DEX0449_022.nt.2	17q25.3	2750-3608	166	DEX0449_022.aa.2
46	DEX0449_023.nt.1	16p13.2	1330-1839	167	DEX0449_023.orf.1
46	DEX0449_023.nt.1	16p13.2	1362-1840	168	DEX0449_023.aa.1
47	DEX0449_023.nt.2	16p13.2	1108-1617	169	DEX0449_023.orf.2
47	DEX0449_023.nt.2	16p13.2	1140-1618	168	DEX0449_023.aa.1
48	DEX0449_024.nt.1	16p13.2	502-1266	170	DEX0449_024.orf.1
48	DEX0449_024.nt.1	16p13.2	590-1034	171	DEX0449_024.aa.1
49	DEX0449_024.nt.2	16p13.2	264-1028	172	DEX0449_024.orf.2
49	DEX0449_024.nt.2	16p13.2	352-796	171	DEX0449_024.aa.1
50	DEX0449_024.nt.3	16p13.2	2-730	173	DEX0449_024.orf.3
50	DEX0449_024.nt.3	16p13.2	55-498	174	DEX0449_024.aa.3
51	DEX0449_024.nt.4	16p13.2	948-1553	175	DEX0449_024.orf.4
51	DEX0449_024.nt.4	16p13.2	1104-1521	176	DEX0449_024.aa.4
52	DEX0449_024.nt.5	16p13.2	232-789	177	DEX0449_024.aa.5
52	DEX0449_024.nt.5	16p13.2	30-824	178	DEX0449_024.orf.5
53	DEX0449_024.nt.6	16p13.2	506-1208	179	DEX0449_024.aa.6
54	DEX0449_025.nt.1	3q25.32	93-569	180	DEX0449_025.aa.1
55	DEX0449_025.nt.2	3q25.32	454-837	181	DEX0449_025.aa.2
56	DEX0449_026.nt.1	5q35.1	90-761	182	DEX0449_026.aa.1

57	DEX0449_026.nt.2	5q35.1	1108-1801	183	DEX0449_026.aa.2
57	DEX0449_026.nt.2	5q35.1	1123-1797	184	DEX0449_026.orf.2
58	DEX0449_027.nt.1	19p13.3	157-3416	185	DEX0449_027.aa.1
58	DEX0449_027.nt.1	19p13.3	3-2006	186	DEX0449_027.orf.1
59	DEX0449_027.nt.2	19p13.3	231-2543	187	DEX0449_027.aa.2
60	DEX0449_027.nt.3	19p13.3	159-2702	188	DEX0449_027.aa.3
61	DEX0449_028.nt.1	8q13.2	152-523	189	DEX0449_028.aa.1
62	DEX0449_029.nt.1	1q21.3	187-1074	190	DEX0449_029.aa.1
63	DEX0449_030.nt.1	20q11.23	1-589	191	DEX0449_030.aa.1
63	DEX0449_030.nt.1	20q11.23	3-584	192	DEX0449_030.orf.1
64	DEX0449_031.nt.1	20q13.33	1827-2226	193	DEX0449_031.aa.1
64	DEX0449_031.nt.1	20q13.33	1602-2198	194	DEX0449_031.orf.1
65	DEX0449_032.nt.1	11q11	133-913	195	DEX0449_032.aa.1
66	DEX0449_032.nt.2	11q11	252-723	196	DEX0449_032.aa.2
66	DEX0449_032.nt.2	11q11	212-715	197	DEX0449_032.orf.2
67	DEX0449_032.nt.3	11q11	134-544	198	DEX0449_032.aa.3
68	DEX0449_032.nt.4	11q11	97-810	199	DEX0449_032.aa.4
68	DEX0449_032.nt.4	11q11	341-808	200	DEX0449_032.orf.4
69	DEX0449_033.nt.1	22q13.31	303-566	201	DEX0449_033.aa.1
70	DEX0449_033.nt.2	22q13.31	43-166	202	DEX0449_033.aa.2
70	DEX0449_033.nt.2	22q13.31	33-212	203	DEX0449_033.orf.2
71	DEX0449_034.nt.1	19p13.2	34-672	204	DEX0449_034.aa.1
72	DEX0449_035.nt.1	3q13.11	1-1378	205	DEX0449_035.aa.1
72	DEX0449_035.nt.1	3q13.11	60-1376	206	DEX0449_035.orf.1
73	DEX0449_036.nt.1	1q42.12	79-473	207	DEX0449_036.aa.1
73	DEX0449_036.nt.1	1q42.12	17-745	208	DEX0449_036.orf.1
74	DEX0449_037.nt.1	7p15.3	625-1236	209	DEX0449_037.aa.1
75	DEX0449_038.nt.1	20q13.13	2-241	210	DEX0449_038.aa.1
76	DEX0449_038.nt.2	20q13.13	1-244	210	DEX0449_038.aa.1
76	DEX0449_038.nt.2	20q13.13	669-920	211	DEX0449_038.orf.2
77	DEX0449_038.nt.3	20q13.13	1-168	212	DEX0449_038.aa.3
78	DEX0449_038.nt.4	20q13.13	1-244	210	DEX0449_038.aa.1
79	DEX0449_039.nt.1	1p34.1	987-2339	213	DEX0449_039.aa.1
80	DEX0449_040.nt.1	6p25.3	272-1282	214	DEX0449_040.aa.1
81	DEX0449_040.nt.2	6p25.3	272-1267	215	DEX0449_040.aa.2
82	DEX0449_040.nt.3	6p25.3	272-1417	216	DEX0449_040.aa.3
83	DEX0449_041.nt.1	19	1892-2665	217	DEX0449_041.aa.1
84	DEX0449_041.nt.2	19	1653-2442	218	DEX0449_041.aa.2

84	DEX0449_041.nt.2	19	1620-2492	219	DEX0449_041.orf.2
85	DEX0449_041.nt.3	19	1891-2593	220	DEX0449_041.aa.3
85	DEX0449_041.nt.3	19	1475-2644	221	DEX0449_041.orf.3
86	DEX0449_042.nt.1	17q11.1	3-332	222	DEX0449_042.aa.1
87	DEX0449_043.nt.1	X;22884428-22906568	263-1033	223	DEX0449_043.aa.1
88	DEX0449_044.nt.1	5p13.1	143-461	224	DEX0449_044.aa.1
88	DEX0449_044.nt.1	5p13.1	125-457	225	DEX0449_044.orf.1
89	DEX0449_044.nt.2	5p13.1	191-506	226	DEX0449_044.aa.2
89	DEX0449_044.nt.2	5p13.1	16-402	227	DEX0449_044.orf.2
90	DEX0449_044.nt.3	5p13.1	3-290	228	DEX0449_044.aa.3
91	DEX0449_044.nt.4	5p13.1	1-168	229	DEX0449_044.aa.4
91	DEX0449_044.nt.4	5p13.1	51-266	230	DEX0449_044.orf.4
92	DEX0449_045.nt.1	19q13.41	5-563	231	DEX0449_045.aa.1
92	DEX0449_045.nt.1	19q13.41	462-1103	232	DEX0449_045.orf.1
93	DEX0449_045.nt.2	16q23.1	5-398	233	DEX0449_045.aa.2
93	DEX0449_045.nt.2	16q23.1	1-396	234	DEX0449_045.orf.2
94	DEX0449_045.nt.3	19q13.41	10-594	235	DEX0449_045.aa.3
95	DEX0449_046.nt.1	X;46271396-46272517	3-347	236	DEX0449_046.aa.1
96	DEX0449_046.nt.2	X;46271094-46272517	2-469	237	DEX0449_046.aa.2
97	DEX0449_047.nt.1	Un_1;3772342-3781437	19-696	238	DEX0449_047.aa.1
97	DEX0449_047.nt.1	Un_1;3772342-3781437	2470-3228	239	DEX0449_047.orf.1
98	DEX0449_047.nt.2	Un_1;3772342-3781437	19-699	238	DEX0449_047.aa.1
98	DEX0449_047.nt.2	Un_1;3772342-3781437	2470-3507	240	DEX0449_047.orf.2
99	DEX0449_048.nt.1	11p15.5	1-277	241	DEX0449_048.aa.1
99	DEX0449_048.nt.1	11p15.5	91-366	242	DEX0449_048.orf.1
100	DEX0449_049.nt.1	3q21.3	34-447	243	DEX0449_049.aa.1
100	DEX0449_049.nt.1	3q21.3	1-444	244	DEX0449_049.orf.1
101	DEX0449_050.nt.1	3q21.3	66-1522	245	DEX0449_050.aa.1
102	DEX0449_050.nt.2	3q21.3	86-1168	246	DEX0449_050.aa.2
103	DEX0449_050.nt.3	3q21.3	66-1345	247	DEX0449_050.aa.3
104	DEX0449_051.nt.1	5q35.3	2-385	248	DEX0449_051.aa.1
105	DEX0449_052.nt.1	13q34	141-1085	249	DEX0449_052.aa.1
106	DEX0449_053.nt.1	7q36.3	674-1099	250	DEX0449_053.aa.1
107	DEX0449_053.nt.2	7q36.3	402-621	251	DEX0449_053.aa.2
107	DEX0449_053.nt.2	7q36.3	622-987	252	DEX0449_053.orf.2
108	DEX0449_054.nt.1	9q34.11	64-194	253	DEX0449_054.aa.1

108	DEX0449_054.nt.1	9q34.11	895-1674	254	DEX0449_054.orf.1
109	DEX0449_054.nt.2	9q34.11	64-194	253	DEX0449_054.aa.1
109	DEX0449_054.nt.2	9q34.11	195-1049	255	DEX0449_054.orf.2
110	DEX0449_055.nt.1	19p13.3	87-1890	256	DEX0449_055.aa.1
110	DEX0449_055.nt.1	19p13.3	28-1887	257	DEX0449_055.orf.1
111	DEX0449_055.nt.2	19p13.3	26-736	258	DEX0449_055.aa.2
112	DEX0449_055.nt.3	19p13.3	87-1890	256	DEX0449_055.aa.1
112	DEX0449_055.nt.3	19p13.3	28-1887	259	DEX0449_055.orf.3

The polypeptides of the present invention were analyzed and the following attributes were identified; specifically, epitopes, post translational modifications, signal peptides and transmembrane domains. Antigenicity (Epitope) prediction was performed through the antigenic module in the EMBOSS package. Rice, P., EMBOSS: The European Molecular Biology Open Software Suite, *Trends in Genetics* 16(6): 276-277 (2000). The antigenic module predicts potentially antigenic regions of a protein sequence, using the method of Kolaskar and Tongaonkar. Kolaskar, AS and Tongaonkar, PC., A semi-empirical method for prediction of antigenic determinants on protein antigens, *FEBS Letters* 276: 172-174 (1990). Examples of post-translational modifications (PTMs) and other motifs of the CSPs of this invention are listed below. In addition, antibodies that specifically bind such post-translational modifications may be useful as a diagnostic or as therapeutic. The PTMs and other motifs were predicted by using the ProSite Dictionary of Proteins Sites and Patterns (Bairoch *et al., Nucleic Acids Res.* 25(1):217-221 (1997)), the following motifs, including PTMs, were predicted for the CSPs of the invention. The signal peptides were detected by using the SignalP 2.0, *see* Nielsen *et al., Protein Engineering* 12, 3-9 (1999). Prediction of transmembrane helices in proteins was performed by the application TMHMM 2.0, "currently the best performing transmembrane prediction program", according to authors (Krogh *et al., Journal of Molecular Biology*, 305(3):567-580, (2001); Moller *et al., Bioinformatics*, 17(7):646-653, (2001); Sonnhammer, *et al., A hidden Markov model for predicting transmembrane helices in protein sequences* in Glasgow, *et al.* Ed. Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology, pages 175-182, Menlo Park, CA, 1998. AAAI Press. The PSORT II program may also be used to predict cellular localizations. Horton et al., *Intelligent Systems for Molecular Biology* 5: 147-152 (1997).

The table below includes the following sequence annotations: Signal peptide presence; TM (number of membrane domain, topology in orientation and position); Amino acid location and antigenic index (location, AI score); PTM and other motifs (type, amino acid residue locations); and functional domains (type, amino acid residue locations).

DEX ID	Sig P	TMHMM	Antigenicity	PTM	Domains
DEX0449_001.aa.1	N	0 - o1-165;	51-59,1.044; 98-105,1.041; 124-130,1.064; 65-70,1.029; 143-162,1.163; 83-96,1.122; 108-116,1.104;	MYRISTYL 9-14; MYRISTYL 44-49; MYRISTYL 12-17; CK2_PHOSPHO_SITE 94-97; MYRISTYL 73-78; MYRISTYL 40-45; MYRISTYL 77-82; PKC_PHOSPHO_SITE 118-120; MYRISTYL 17-22; PKC_PHOSPHO_SITE 107-109; CK2_PHOSPHO_SITE 87-90; MYRISTYL 92-97; MYRISTYL 16-21;	RRM 82-152; rrm 84-154; RRM 83-155; GLY_RICH 1-82;
DEX0449_001.orf.1	N	0 - o1-164;		MYRISTYL 3-8; MYRISTYL 72-77; MYRISTYL 15-20; MYRISTYL 91-96; PKC_PHOSPHO_SITE 106-108; CK2_PHOSPHO_SITE 86-89; MYRISTYL 39-44; MYRISTYL 16-21; MYRISTYL 43-48; MYRISTYL 11-16; CK2_PHOSPHO_SITE 93-96; MYRISTYL 76-81; PKC_PHOSPHO_SITE 117-119;	RRM 82-154; RRM 81-151; GLY_RICH 1-81; rrm 83-153;
DEX0449_002.aa.1	N	0 - o1-256;	134-156,1.159; 116-125,1.111; 24-34,1.087; 186-196,1.13; 71-101,1.22; 203-226,1.227; 164-180,1.143; 237-245,1.093;	PKC_PHOSPHO_SITE 71-73; MYRISTYL 199-204; CK2_PHOSPHO_SITE 128-131; MYRISTYL 49-54; MYRISTYL 233-238; MYRISTYL 134-139; PKC_PHOSPHO_SITE 34-36; MYRISTYL 161-166;	PRO_RICH 80-129;
DEX0449_00	Y	0 - o1-	158-	MYRISTYL 176-181;	PRO_RICH 122-171;

2.orf.1		250;	167,1.111; 66-76,1.087; 113-143,1.22; 176-198,1.159; 9-36,1.178; 206-222,1.143; 228-238,1.13;	PKC_PHOSPHO_SITE 76-78; MYRISTYL 91-96; PKC_PHOSPHO_SITE 4-6; MYRISTYL 241-246; MYRISTYL 203-208; CK2_PHOSPHO_SITE 170-173; PKC_PHOSPHO_SITE 3-5; PKC_PHOSPHO_SITE 113-115;	
DEX0449_003.aa.1	N	0 - o1-406;	82-97,1.113; 201-207,1.14; 59-75,1.139; 124-130,1.123; 213-226,1.126; 142-174,1.207; 356-375,1.115; 251-278,1.154; 388-403,1.127; 334-340,1.063; 323-332,1.205; 182-191,1.186; 105-121,1.136; 4-11,1.111; 239-248,1.148; 289-300,1.104; 20-26,1.083;	MYRISTYL 335-340; CK2_PHOSPHO_SITE 30-33; CK2_PHOSPHO_SITE 49-52; PKC_PHOSPHO_SITE 120-122; CK2_PHOSPHO_SITE 89-92; MYRISTYL 198-203; PKC_PHOSPHO_SITE 13-15; PKC_PHOSPHO_SITE 49-51; CK2_PHOSPHO_SITE 88-91; PKC_PHOSPHO_SITE 79-81; PKC_PHOSPHO_SITE 344-346; MYRISTYL 189-194; ASN_GLYCOSYLATION 379-382;	Glyco_hydro_85 131-385;
DEX0449_003.aa.3	N	0 - o1-525;	450-457,1.192; 411-418,1.091; 124-130,1.123; 347-373,1.136; 20-26,1.083; 490-502,1.207; 182-191,1.186; 289-300,1.104; 213-	PKC_PHOSPHO_SITE 79-81; MYRISTYL 198-203; CK2_PHOSPHO_SITE 88-91; MYRISTYL 189-194; PKC_PHOSPHO_SITE 344-346; PKC_PHOSPHO_SITE 120-122; PKC_PHOSPHO_SITE 13-15; MYRISTYL 335-340; MYRISTYL 486-491; CK2_PHOSPHO_SITE 49-52;	BRCT 337-383; Glyco_hydro_85 131-414;

			226,1.126; 105-121,1.136; 239-248,1.148; 251-278,1.154; 142-174,1.207; 82-97,1.113; 59-75,1.139; 334-340,1.063; 462-483,1.166; 422-438,1.104; 323-332,1.205; 201-207,1.14; 4-11,1.111; 387-408,1.195;	CK2_PHOSPHO_SITE 428-431; PKC_PHOSPHO_SITE 49-51; CK2_PHOSPHO_SITE 30-33; MYRISTYL 416-421; CK2_PHOSPHO_SITE 89-92;	
DEX0449_003.aa.4	N	0 - o1-327;	11-18,1.089; 270-289,1.165; 90-97,1.156; 134-143,1.053; 29-36,1.1; 60-86,1.143; 44-57,1.084; 317-324,1.169; 223-250,1.179; 120-132,1.108; 146-171,1.237; 298-314,1.155; 182-214,1.168; 102-115,1.1;	CK2_PHOSPHO_SITE 175-178; MYRISTYL 37-42; CK2_PHOSPHO_SITE 216-219; MYRISTYL 272-277; PKC_PHOSPHO_SITE 182-184; PKC_PHOSPHO_SITE 218-220; CK2_PHOSPHO_SITE 218-221; PKC_PHOSPHO_SITE 122-124; PKC_PHOSPHO_SITE 156-158; CK2_PHOSPHO_SITE 98-101; PKC_PHOSPHO_SITE 118-120; MYRISTYL 9-14;	
DEX0449_003.orf.4	N	0 - o1-384;	119-143,1.143; 355-371,1.155; 280-307,1.179; 177-189,1.108; 95-112,1.089; 327-346,1.165; 374-	PKC_PHOSPHO_SITE 90-92; PKC_PHOSPHO_SITE 8-10; CK2_PHOSPHO_SITE 273-276; PKC_PHOSPHO_SITE 175-177; CK2_PHOSPHO_SITE 232-235; PKC_PHOSPHO_SITE 179-181; MYRISTYL 329-334;	

			381,1.169; 85-90,1.061; 239-271,1.168; 191-200,1.053; 159-172,1.1; 203-228,1.237; 147-154,1.156; 4-64,1.135;	CK2_PHOSPHO_SITE 155-158; CK2_PHOSPHO_SITE 275-278; MYRISTYL 94-99; PKC_PHOSPHO_SITE 239-241; MYRISTYL 81-86; PKC_PHOSPHO_SITE 275-277; CK2_PHOSPHO_SITE 90-93; MYRISTYL 47-52; PKC_PHOSPHO_SITE 213-215;	
DEX0449_003.orf.6	N	0 - o1-450;	147-154,1.156; 95-112,1.089; 255-269,1.178; 159-172,1.1; 421-437,1.155; 292-337,1.244; 4-64,1.135; 440-447,1.169; 239-250,1.08; 393-412,1.165; 177-189,1.108; 191-200,1.053; 85-90,1.061; 119-143,1.143; 346-373,1.179; 271-278,1.063; 203-228,1.237;	CK2_PHOSPHO_SITE 155-158; PKC_PHOSPHO_SITE 90-92; CK2_PHOSPHO_SITE 90-93; CK2_PHOSPHO_SITE 232-235; PKC_PHOSPHO_SITE 8-10; MYRISTYL 47-52; MYRISTYL 252-257; MYRISTYL 287-292; PKC_PHOSPHO_SITE 179-181; MYRISTYL 395-400; PKC_PHOSPHO_SITE 341-343; CK2_PHOSPHO_SITE 341-344; PKC_PHOSPHO_SITE 213-215; CK2_PHOSPHO_SITE 339-342; PKC_PHOSPHO_SITE 239-241; MYRISTYL 81-86; MYRISTYL 94-99; PKC_PHOSPHO_SITE 175-177;	
DEX0449_003.aa.7	N	0 - o1-302;	82-97,1.113; 292-299,1.169; 273-289,1.155; 20-26,1.083; 124-130,1.123; 105-121,1.136; 239-248,1.148; 182-	PKC_PHOSPHO_SITE 79-81; CK2_PHOSPHO_SITE 30-33; MYRISTYL 189-194; PKC_PHOSPHO_SITE 120-122; MYRISTYL 198-203; CK2_PHOSPHO_SITE 49-52; PKC_PHOSPHO_SITE 49-51; CK2 PHOSPHO SITE	Glyco_hydro_85 131-297;

			191,1.186; 251-264,1.138; 201-207,1.14; 213-226,1.126; 59-75,1.139; 4-11,1.111; 142-174,1.207;	88-91; PKC_PHOSPHO_SITE 13-15; CK2_PHOSPHO_SITE 89-92;	
DEX0449_004.aa.1	Y	1 - o1-62;tm63-85;i86-162;	4-19,1.198; 55-67,1.104; 22-48,1.14; 69-116,1.236;	PKC_PHOSPHO_SITE 87-89; CK2_PHOSPHO_SITE 98-101; CK2_PHOSPHO_SITE 135-138; MYRISTYL 128-133;	Sterol_desat 1-129; SUR2_DOMAIN 18-126;
DEX0449_005.aa.1	N	0 - o1-206;	186-197,1.16; 36-45,1.101; 97-119,1.199; 163-173,1.076; 142-147,1.042; 8-31,1.142; 58-76,1.16;	CK2_PHOSPHO_SITE 100-103; PKC_PHOSPHO_SITE 7-9; CK2_PHOSPHO_SITE 42-45; PKC_PHOSPHO_SITE 194-196; PKC_PHOSPHO_SITE 42-44; MYRISTYL 75-80; PKC_PHOSPHO_SITE 183-185;	
DEX0449_005.orf.1	N	0 - o1-380;	156-178,1.199; 349-355,1.062; 243-259,1.168; 117-135,1.16; 95-104,1.101; 4-10,1.175; 275-293,1.148; 360-371,1.088; 43-54,1.098; 315-322,1.094; 217-235,1.131; 296-304,1.127; 17-31,1.101; 261-267,1.072; 201-206,1.042; 67-90,1.142;	MYRISTYL 261-266; PKC_PHOSPHO_SITE 101-103; CK2_PHOSPHO_SITE 40-43; MYRISTYL 52-57; PKC_PHOSPHO_SITE 310-312; PKC_PHOSPHO_SITE 332-334; CK2_PHOSPHO_SITE 333-336; CK2_PHOSPHO_SITE 101-104; PKC_PHOSPHO_SITE 66-68; CK2_PHOSPHO_SITE 310-313; CK2_PHOSPHO_SITE 302-305; MYRISTYL 134-139; CK2_PHOSPHO_SITE 159-162; PKC_PHOSPHO_SITE 9-11; MYRISTYL 221-226;	

DEX0449_005.orf.2	N	0 - o1-380;	217-235,1.131; 296-304,1.127; 156-178,1.199; 95-104,1.101; 349-355,1.062; 67-90,1.142; 17-31,1.101; 201-206,1.042; 243-259,1.168; 261-267,1.072; 43-54,1.098; 315-322,1.094; 4-10,1.175; 275-293,1.148; 360-371,1.088; 117-135,1.16;	PKC_PHOSPHO_SITE 101-103; PKC_PHOSPHO_SITE 66-68; CK2_PHOSPHO_SITE 310-313; CK2_PHOSPHO_SITE 302-305; MYRISTYL 261-266; PKC_PHOSPHO_SITE 310-312; MYRISTYL 134-139; CK2_PHOSPHO_SITE 159-162; MYRISTYL 52-57; PKC_PHOSPHO_SITE 9-11; CK2_PHOSPHO_SITE 101-104; CK2_PHOSPHO_SITE 40-43; CK2_PHOSPHO_SITE 333-336; PKC_PHOSPHO_SITE 332-334; MYRISTYL 221-226;	
DEX0449_005.aa.3	N	0 - o1-334;	94-105,1.164; 46-52,1.071; 114-124,1.143; 270-275,1.042; 225-247,1.199; 314-325,1.16; 186-204,1.16; 291-301,1.076; 136-159,1.142; 10-16,1.095; 18-37,1.212; 164-173,1.101; 76-86,1.121; 59-65,1.085;	CAMP_PHOSPHO_SITE 73-76; MYRISTYL 95-100; PKC_PHOSPHO_SITE 135-137; CK2_PHOSPHO_SITE 170-173; MYRISTYL 85-90; PKC_PHOSPHO_SITE 170-172; PKC_PHOSPHO_SITE 322-324; MYRISTYL 71-76; MYRISTYL 51-56; MYRISTYL 31-36; PKC_PHOSPHO_SITE 311-313; PKC_PHOSPHO_SITE 28-30; MYRISTYL 82-87; MYRISTYL 46-51; MYRISTYL 203-208; CK2_PHOSPHO_SITE 228-231;	
DEX0449_005.orf.3	N	0 - o1-327;	307-318,1.088; 64-82,1.16; 243-251,1.127; 262-269,1.094;	MYRISTYL 168-173; CK2_PHOSPHO_SITE 257-260; MYRISTYL 81-86; PKC_PHOSPHO_SITE 257-259; CK2_PHOSPHO_SITE	

			164-182,1.131; 148-153,1.042; 42-51,1.101; 296-302,1.062; 222-240,1.148; 14-37,1.142; 208-214,1.072; 190-206,1.168; 103-125,1.199;	280-283; MYRISTYL 208-213; CK2_PHOSPHO_SITE 48-51; CK2_PHOSPHO_SITE 106-109; CK2_PHOSPHO_SITE 3-6; CK2_PHOSPHO_SITE 249-252; PKC_PHOSPHO_SITE 48-50; PKC_PHOSPHO_SITE 279-281; PKC_PHOSPHO_SITE 13-15;	
DEX0449_005.aa.4	N	0 - o1-161;	36-45,1.101; 142-147,1.042; 58-76,1.16; 97-119,1.199; 8-31,1.142;	PKC_PHOSPHO_SITE 42-44; ASN_GLYCOSYLATION 154-157; MYRISTYL 75-80; CK2_PHOSPHO_SITE 42-45; PKC_PHOSPHO_SITE 156-158; PKC_PHOSPHO_SITE 7-9; CK2_PHOSPHO_SITE 100-103;	
DEX0449_005.orf.6	N	0 - o1-326;	156-178,1.199; 217-235,1.131; 43-54,1.098; 67-90,1.142; 296-314,1.123; 261-267,1.072; 117-135,1.16; 275-293,1.148; 201-206,1.042; 243-259,1.168; 17-31,1.101; 4-10,1.175; 95-104,1.101;	MYRISTYL 261-266; PKC_PHOSPHO_SITE 9-11; MYRISTYL 221-226; PKC_PHOSPHO_SITE 66-68; CK2_PHOSPHO_SITE 101-104; MYRISTYL 134-139; MYRISTYL 52-57; PKC_PHOSPHO_SITE 101-103; CK2_PHOSPHO_SITE 40-43; CK2_PHOSPHO_SITE 159-162;	
DEX0449_006.aa.1	N	0 - o1-216;	23-80,1.213; 88-99,1.097; 123-140,1.268; 9-15,1.084; 109-116,1.12; 142-	CK2_PHOSPHO_SITE 89-92; PKC_PHOSPHO_SITE 99-101; MYRISTYL 84-89; CAMP_PHOSPHO_SITE 79-82;	PRO_RICH 31-209;

			207,1.19;		
DEX0449_007.aa.1	Y	0 - o1-108;	4-10,1.067; 46-52,1.055; 12-30,1.269; 59-105,1.18; 39-44,1.054;	PKC_PHOSPHO_SITE 36-38; PKC_PHOSPHO_SITE 29-31;	SMALLCYTKCXC 75-83; SMALLCYTKCXC 90-108;
DEX0449_007.orf.1	Y	0 - o1-142;	67-73,1.058; 46-58,1.043; 93-139,1.18; 4-30,1.163; 37-44,1.048;	PKC_PHOSPHO_SITE 10-12; MYRISTYL 7-12;	SMALLCYTKCXC 124-142; INTERLEUKIN8 125-142; INTERLEUKIN8 102-125; SMALLCYTKCXC 109-117;
DEX0449_008.aa.1	N	0 - o1-482;	298-306,1.134; 176-183,1.113; 16-33,1.107; 325-341,1.104; 38-45,1.113; 373-394,1.127; 285-294,1.256; 257-275,1.079; 354-361,1.095; 230-254,1.098; 52-58,1.067; 121-128,1.113; 7-14,1.136; 110-119,1.111; 62-101,1.184;	MYRISTYL 44-49; ASN_GLYCOSYLATION 317-320; PKC_PHOSPHO_SITE 48-50; ASN_GLYCOSYLATION 285-288; MYRISTYL 43-48; CK2_PHOSPHO_SITE 141-144; TYR_PHOSPHO_SITE 50-56; MYRISTYL 96-101; ASN_GLYCOSYLATION 69-72; MYRISTYL 258-263; MYRISTYL 471-476; ASN_GLYCOSYLATION 37-40; PKC_PHOSPHO_SITE 250-252; CK2_PHOSPHO_SITE 204-207; MYRISTYL 435-440; CK2_PHOSPHO_SITE 85-88; MYRISTYL 465-470; CK2_PHOSPHO_SITE 192-195; PKC_PHOSPHO_SITE 466-468; MYRISTYL 118-123; MYRISTYL 439-444; CK2_PHOSPHO_SITE 149-152;	RRM 332-405; RRM 331-409; NTF2 11-133; GLY_RICH 419-479; NTF2_DOMAIN 11-133; rrm 333-404; sp_Q9UPA1_Q9UPA1_HUMAN 11-132; GLU_RICH 134-223;
DEX0449_008.orf.1	N	0 - o1-392;	4-12,1.235; 264-271,1.095; 283-304,1.127; 235-251,1.104; 86-93,1.113; 37-45,1.123; 140-165.1.098:	CK2_PHOSPHO_SITE 59-62; RGD 221-223; MYRISTYL 168-173; CK2_PHOSPHO_SITE 114-117; CK2_PHOSPHO_SITE 102-105; MYRISTYL 345-350; PKC_PHOSPHO_SITE 376-378:	RRM 242-315; HUDSXLRNA 246-261; HUDSXLRNA 261-273; rrm 243-314; PRICHEXTENSN 101-122; PRICHEXTENSN 194-219; GLU_RICH 52-133; PRICHEXTENSN 164-181; GLY_RICH 329-389: RRM 241-319:

			16-33,1.144; 167-201,1.142;	PKC_PHOSPHO_SITE 160-162; MYRISTYL 375-380; ASN_GLYCOSYLATION 227-230; MYRISTYL 349-354; CK2_PHOSPHO_SITE 51-54; MYRISTYL 381-386;	PRICHEXTENSN 136-152; PRO_RICH 138-219;
DEX0449_009.aa.1	N	0 - o1-316;	159-175,1.104; 233-239,1.056; 94-107,1.098; 207-228,1.127; 109-128,1.075; 30-37,1.065; 5-15,1.225; 188-195,1.095; 41-50,1.113; 135-140,0.986;	CK2_PHOSPHO_SITE 71-74; CK2_PHOSPHO_SITE 16-19; MYRISTYL 269-274; ASN_GLYCOSYLATION 151-154; MYRISTYL 305-310; PKC_PHOSPHO_SITE 300-302; CK2_PHOSPHO_SITE 14-17; MYRISTYL 299-304; MYRISTYL 273-278; RGD 145-147; CK2_PHOSPHO_SITE 59-62;	RRM 166-239; rrm 167-238; HUDSXLRNA 170-185; HUDSXLRNA 185-197; GLU_RICH 17-90; RRM 165-243; PRO_RICH 95-143; GLY_RICH 253-313;
DEX0449_009.orf.1	N	0 - o1-314;		MYRISTYL 267-272; MYRISTYL 271-276; ASN_GLYCOSYLATION 149-152; MYRISTYL 2-7; ASN_GLYCOSYLATION 4-7; RGD 143-145; PKC_PHOSPHO_SITE 298-300; CK2_PHOSPHO_SITE 57-60; MYRISTYL 297-302; CK2_PHOSPHO_SITE 14-17; MYRISTYL 303-308; CK2_PHOSPHO_SITE 12-15; CK2_PHOSPHO_SITE 69-72;	HUDSXLRNA 183-195; GLU_RICH 15-88; PRO_RICH 93-141; RRM 164-237; rrm 165-236; GLY_RICH 251-311; HUDSXLRNA 168-183; RRM 163-241;
DEX0449_009.aa.2	N	0 - i1-169;	41-48,1.095; 125-130,1.013; 11-28,1.104; 86-92,1.056; 55-82,1.127;	MYRISTYL 126-131; MYRISTYL 122-127; MYRISTYL 152-157; MYRISTYL 158-163; ASN_GLYCOSYLATION 4-7; PKC_PHOSPHO_SITE 153-155;	RRM 19-92; HUDSXLRNA 23-38; GLY_RICH 106-166; RRM 18-96; HUDSXLRNA 38-50; rrm 20-91;
DEX0449_010.aa.1	N	0 - i1-147;	85-99,1.144; 9-23,1.113; 66-73,1.073; 51-59,1.107; 106-	PKC_PHOSPHO_SITE 43-45; PKC_PHOSPHO_SITE 8-10; MYRISTYL 112-117;	Ribosomal_S17e 1-122; RIBOSOMAL_S17E 41-56;

			112,1.085; 114-134,1.13; 31-44,1.124;	PKC_PHOSPHO_SITE 70-72; TYR_PHOSPHO_SITE 14-21; PKC_PHOSPHO_SITE 30-32; CK2_PHOSPHO_SITE 89-92; PKC_PHOSPHO_SITE 142-144;	
DEX0449_010.orf.1	Y	0 - o1-166;		PKC_PHOSPHO_SITE 27-29; PKC_PHOSPHO_SITE 89-91; PKC_PHOSPHO_SITE 62-64; PKC_PHOSPHO_SITE 161-163; TYR_PHOSPHO_SITE 33-40; PKC_PHOSPHO_SITE 49-51; CK2_PHOSPHO_SITE 108-111; MYRISTYL 131-136; PKC_PHOSPHO_SITE 132-134;	Ribosomal_S17e 20-141; RIBOSOMAL_S17E 60-75;
DEX0449_011.aa.1	N	0 - o1-254;	90-105,1.105; 221-232,1.115; 213-219,1.125; 15-32,1.096; 160-173,1.13; 239-247,1.187; 77-87,1.09;	CK2_PHOSPHO_SITE 209-212; CAMP_PHOSPHO_SITE 7-10; CK2_PHOSPHO_SITE 35-38; CK2_PHOSPHO_SITE 233-236; ASN_GLYCOSYLATION 192-195; PKC_PHOSPHO_SITE 35-37; ASN_GLYCOSYLATION 223-226; CK2_PHOSPHO_SITE 11-14; PKC_PHOSPHO_SITE 67-69; CK2_PHOSPHO_SITE 107-110; MYRISTYL 64-69; CK2_PHOSPHO_SITE 10-13; PKC_PHOSPHO_SITE 25-27; CK2_PHOSPHO_SITE 123-126; PKC_PHOSPHO_SITE 92-94; ASN_GLYCOSYLATION 65-68; TYR PHOSPHO SITE	TFIIS 216-251; ZnF_C2C2 214-253; TFS2M 91-192; TFSII 1-254; TFIIS 214-252;

				134-140; CK2_PHOSPHO_SITE 177-180; CK2_PHOSPHO_SITE 88-91; MYRISTYL 207-212;	
DEX0449_011.orf.1	Y	0 - o1-302;	208-221,1.13; 261-267,1.125; 4-20,1.264; 125-134,1.09; 138-148,1.105; 63-76,1.096; 25-32,1.092; 269-280,1.115; 38-45,1.077; 287-294,1.187;	CAMP_PHOSPHO_SITE 55-58; ASN_GLYCOSYLATION 271-274; CK2_PHOSPHO_SITE 59-62; PKC_PHOSPHO_SITE 140-142; CK2_PHOSPHO_SITE 83-86; TYR_PHOSPHO_SITE 182-188; CK2_PHOSPHO_SITE 281-284; CK2_PHOSPHO_SITE 58-61; MYRISTYL 255-260; CK2_PHOSPHO_SITE 136-139; ASN_GLYCOSYLATION 240-243; CK2_PHOSPHO_SITE 171-174; PKC_PHOSPHO_SITE 44-46; CK2_PHOSPHO_SITE 225-228; PKC_PHOSPHO_SITE 115-117; CK2_PHOSPHO_SITE 257-260; PKC_PHOSPHO_SITE 83-85; MYRISTYL 112-117; ASN_GLYCOSYLATION 113-116; PKC_PHOSPHO_SITE 73-75; MYRISTYL 48-53; CK2_PHOSPHO_SITE 155-158;	TFSII 2-302; TFIIS 264-299; TFS2M 139-240; TFS2N 12-80; ZnF_C2C2 262-301; TFIIS 262-300;
DEX0449_012.aa.1	N	0 - o1-225;	92-98,1.091; 159-168,1.118; 34-43,1.152; 132-138,1.052; 113-125,1.136; 63-71,1.115; 4-26,1.152; 73-89,1.107; 100-	MYRISTYL 110-115; CK2_PHOSPHO_SITE 114-117; CK2_PHOSPHO_SITE 3-6; CK2_PHOSPHO_SITE 182-185; PKC_PHOSPHO_SITE 63-65; MYRISTYL 39-44; PKC_PHOSPHO_SITE 48-50;	HAD-SF-IB 16-190; serB 1-220; Hydrolase 14-200;

			110,1.184; 186-195,1.089; 54-61,1.121; 213-222,1.101;		
DEX0449_013.aa.1	N	0 - o1-249;	33-43,1.16; 210-218,1.114; 198-208,1.067; 133-155,1.217; 163-191,1.174; 74-80,1.077; 113-123,1.179; 232-246,1.117; 83-106,1.227; 54-71,1.119; 21-28,1.053; 125-131,1.082;	PKC_PHOSPHO_SITE 16-18; CK2_PHOSPHO_SITE 216-219; MYRISTYL 41-46; PKC_PHOSPHO_SITE 204-206; MYRISTYL 212-217; PKC_PHOSPHO_SITE 15-17; MYRISTYL 33-38; CK2_PHOSPHO_SITE 47-50; TYR_PHOSPHO_SITE 61-68; CK2_PHOSPHO_SITE 162-165; MYRISTYL 138-143;	
DEX0449_014.aa.1	N	0 - o1-113;	31-45,1.147; 57-80,1.147;	PKC_PHOSPHO_SITE 23-25; CK2_PHOSPHO_SITE 107-110; PKC_PHOSPHO_SITE 17-19; CK2_PHOSPHO_SITE 98-101; PKC_PHOSPHO_SITE 8-10; CK2_PHOSPHO_SITE 54-57; ASN_GLYCOSYLATION 52-55; CK2_PHOSPHO_SITE 17-20; ASN_GLYCOSYLATION 14-17;	RCC1_2 66-76;
DEX0449_015.aa.1	N	0 - o1-102;	83-89,1.072; 17-38,1.144; 43-51,1.06; 65-78,1.148; 91-99,1.076;	PKC_PHOSPHO_SITE 59-61;	
DEX0449_016.aa.1	N	0 - o1-412;	362-391,1.147; 263-294,1.164; 297-318,1.186; 351-359,1.147; 329-	MYRISTYL 185-190; MYRISTYL 226-231; CK2_PHOSPHO_SITE 134-137; PKC_PHOSPHO_SITE 398-400; MYRISTYL 38-43; MYRISTYL 383-388; CK2 PHOSPHO SITE	CYTOCHROME_P450 348-357; PRICHEXTENSN 46-67; EP450I 270-288; P450 355-366; PRICHEXTENSN 204-229; EP450I 345-355; EP450I 173-190; P450 271-282;

			349,1.091; 403-409,1.179; 44-70,1.113; 245-254,1.08; 191-201,1.112; 233-243,1.061; 84-125,1.211; 171-182,1.135; 23-36,1.193; 5-21,1.155; 127-159,1.183;	290-293; AMIDATION 350-353; CK2_PHOSPHO_SITE 151-154; PKC_PHOSPHO_SITE 257-259; MYRISTYL 42-47; MYRISTYL 222-227; PKC_PHOSPHO_SITE 2-4;	EP450I 193-219; P450 184-201; p450 3-408; EP450I 355-378; EP450I 310-334; P450 346-355;
DEX0449_016.orf.1	N	0 - o1-203;	36-45,1.08; 194-200,1.179; 22-34,1.085; 88-109,1.186; 8-14,1.041; 153-182,1.147; 142-150,1.147; 120-140,1.091; 54-85,1.164;	PKC_PHOSPHO_SITE 48-50; PKC_PHOSPHO_SITE 15-17; AMIDATION 141-144; CK2_PHOSPHO_SITE 81-84; CK2_PHOSPHO_SITE 15-18; MYRISTYL 174-179; PKC_PHOSPHO_SITE 189-191;	EP450I 18-44; P450 27-40; BP450 146-157; P450 146-157; EP450I 61-79; BP450 137-146; P450 62-73; CYTOCHROME_P450 139-148; EP450IV 57-73; EP450IV 146-164; EP450I 146-169; EP450IV 130-146; EP450I 136-146; EP450I 101-125; BP450 79-106; EP450IV 106-124; P450 137-146; BP450 62-73;
DEX0449_017.aa.1	Y	0 - i1-116;	74-87,1.16; 99-108,1.095; 4-31,1.198; 48-71,1.145;	ASN_GLYCOSYLATION 94-97; ASN_GLYCOSYLATION 64-67; PKC_PHOSPHO_SITE 98-100;	
DEX0449_017.orf.1	N	0 - o1-141;	126-138,1.16; 25-30,1.033; 100-123,1.145; 38-44,1.04; 73-80,1.087; 87-93,1.083; 11-21,1.245; 58-68,1.175;	PKC_PHOSPHO_SITE 43-45; RGD 58-60; ASN_GLYCOSYLATION 116-119; AMIDATION 21-24; PKC_PHOSPHO_SITE 31-33; MYRISTYL 92-97; MYRISTYL 25-30; MYRISTYL 21-26;	
DEX0449_018.aa.1	N	0 - o1-426;	5-43,1.154; 106-116,1.117; 90-95,1.044; 356-363,1.107; 224-231,1.101;	CAMP_PHOSPHO_SITE 329-332; CK2_PHOSPHO_SITE 377-380; PKC_PHOSPHO_SITE 229-231; PKC_PHOSPHO_SITE 328-330;	ZINC_FINGER_C2H2_1 74-96;

			322-328,1.077; 70-80,1.145; 293-299,1.08; 205-221,1.197; 172-187,1.186; 192-199,1.115; 271-282,1.08; 154-169,1.132; 401-418,1.128; 59-65,1.037; 45-50,1.049; 333-351,1.138;	CK2_PHOSPHO_SITE 135-138; CK2_PHOSPHO_SITE 75-78; TYR_PHOSPHO_SITE 264-272; CK2_PHOSPHO_SITE 120-123; CK2_PHOSPHO_SITE 137-140; PKC_PHOSPHO_SITE 368-370; CAMP_PHOSPHO_SITE 241-244; CK2_PHOSPHO_SITE 128-131; MYRISTYL 129-134; RGD 316-318; PKC_PHOSPHO_SITE 69-71; CK2_PHOSPHO_SITE 111-114; PKC_PHOSPHO_SITE 301-303; PKC_PHOSPHO_SITE 398-400; MYRISTYL 243-248; CK2_PHOSPHO_SITE 332-335; PKC_PHOSPHO_SITE 237-239; MYRISTYL 247-252; CAMP_PHOSPHO_SITE 374-377; CK2_PHOSPHO_SITE 15-18; CK2_PHOSPHO_SITE 368-371;	
DEX0449_018.aa.4	N	0 - o1-370;	237-243,1.08; 215-226,1.08; 345-362,1.128; 168-175,1.101; 266-272,1.077; 98-113,1.132; 116-131,1.186; 50-60,1.117; 5-39,1.125; 149-165,1.197; 136-143,1.115; 300-	TYR_PHOSPHO_SITE 208-216; CK2_PHOSPHO_SITE 276-279; CK2_PHOSPHO_SITE 72-75; CK2_PHOSPHO_SITE 64-67; PKC_PHOSPHO_SITE 342-344; CAMP_PHOSPHO_SITE 185-188; CK2_PHOSPHO_SITE 79-82; PKC_PHOSPHO_SITE 181-183; CK2_PHOSPHO_SITE 55-58; PKC_PHOSPHO_SITE 312-314; PKC PHOSPHO SITE	

			307,1.107; 277-295,1.138;	30-32; CK2_PHOSPHO_SITE 81-84; MYRISTYL 73-78; CK2_PHOSPHO_SITE 30-33; MYRISTYL 191-196; PKC_PHOSPHO_SITE 245-247; CK2_PHOSPHO_SITE 321-324; CAMP_PHOSPHO_SITE 273-276; CK2_PHOSPHO_SITE 312-315; PKC_PHOSPHO_SITE 272-274; RGD 260-262; CK2_PHOSPHO_SITE 15-18; MYRISTYL 187-192; PKC_PHOSPHO_SITE 173-175; CAMP_PHOSPHO_SITE 318-321;	
DEX0449_018.aa.5	N	0 - o1-208;	70-80,1.145; 45-50,1.049; 59-65,1.037; 90-95,1.044; 154-169,1.132; 172-205,1.186; 106-116,1.117; 5-43,1.154;	CK2_PHOSPHO_SITE 111-114; CK2_PHOSPHO_SITE 120-123; PKC_PHOSPHO_SITE 69-71; CK2_PHOSPHO_SITE 135-138; MYRISTYL 129-134; CK2_PHOSPHO_SITE 75-78; CK2_PHOSPHO_SITE 128-131; CK2_PHOSPHO_SITE 15-18; CK2_PHOSPHO_SITE 137-140;	ZINC_FINGER_C2H2_1 74-96;
DEX0449_018.orf.7	N	0 - o1-209;	115-131,1.129; 17-39,1.155; 57-80,1.144; 192-203,1.2; 181-189,1.123; 100-106,1.067; 140-178,1.174; 4-13,1.164;	MYRISTYL 98-103; CK2_PHOSPHO_SITE 50-53; CK2_PHOSPHO_SITE 107-110; CK2_PHOSPHO_SITE 162-165; MYRISTYL 203-208; CK2_PHOSPHO_SITE 48-51;	ank 179-209; ANK_REPEAT 179-209; ANKYRIN 192-204; ANK_REP_REGION 179-209; ANKYRIN 180-192; ANK 138-168; ANK 179-208; ank 138-171;
DEX0449_018.aa.7	N	0 - o1-125;	70-80,1.145; 90-95,1.044; 45-50,1.049; 5-43,1.154; 59-65,1.037;	CK2_PHOSPHO_SITE 121-124; CK2_PHOSPHO_SITE 111-114; CK2_PHOSPHO_SITE	ZINC_FINGER_C2H2_1 74-96;

			106-116,1.117;	15-18; CK2_PHOSPHO_SITE 120-123; PKC_PHOSPHO_SITE 69-71; CK2_PHOSPHO_SITE 75-78;	
DEX0449_019.orf.1	N	0 - o1-191;	45-58,1.139; 90-111,1.126; 63-86,1.209; 178-184,1.058; 115-135,1.185; 10-37,1.175; 140-154,1.113;	PKC_PHOSPHO_SITE 165-167; CK2_PHOSPHO_SITE 38-41; PKC_PHOSPHO_SITE 184-186; PKC_PHOSPHO_SITE 183-185; PKC_PHOSPHO_SITE 57-59; PKC_PHOSPHO_SITE 89-91; MYRISTYL 15-20; MYRISTYL 112-117; PKC_PHOSPHO_SITE 84-86;	
DEX0449_019.aa.1	Y	0 - o1-130;	101-115,1.14; 117-127,1.111; 37-59,1.165; 13-30,1.165; 66-71,1.085; 4-11,1.144; 75-81,1.067;	CK2_PHOSPHO_SITE 79-82; AMIDATION 32-35; ASN_GLYCOSYLATION 63-66; MYRISTYL 38-43; ASN_GLYCOSYLATION 75-78; MYRISTYL 7-12; PKC_PHOSPHO_SITE 59-61; MYRISTYL 2-7;	
DEX0449_020.orf.1	N	0 - o1-340;	216-223,1.099; 183-191,1.146; 127-173,1.255; 269-275,1.09; 112-123,1.195; 315-321,1.08; 79-102,1.145; 226-252,1.213; 286-302,1.253; 23-29,1.075; 194-213,1.151; 259-264,1.066; 326-	CK2_PHOSPHO_SITE 223-226; MYRISTYL 285-290; PKC_PHOSPHO_SITE 267-269; TYR_PHOSPHO_SITE 322-329; PKC_PHOSPHO_SITE 17-19; CAMP_PHOSPHO_SITE 18-21; MYRISTYL 123-128; CAMP_PHOSPHO_SITE 19-22; CK2_PHOSPHO_SITE 255-258; MYRISTYL 171-176; PKC_PHOSPHO_SITE 21-23; CK2_PHOSPHO_SITE 248-251;	aminotran_4 26-325; sp_Q9BTB6_Q9BTB6_HUMAN 192-296; AA_TRANSFER_CLASS_4 212-246;

			332,1.042; 62-68,1.078;		
DEX0449_020.aa.1	N	0 - o1-306;	281-287,1.08; 28-34,1.078; 235-241,1.09; 78-89,1.195; 192-218,1.213; 225-230,1.066; 93-139,1.255; 45-68,1.145; 292-298,1.042; 149-157,1.146; 160-179,1.151; 182-189,1.099; 252-268,1.253;	CK2_PHOSPHO_SITE 189-192; PKC_PHOSPHO_SITE 233-235; TYR_PHOSPHO_SITE 288-295; MYRISTYL 251-256; CK2_PHOSPHO_SITE 221-224; MYRISTYL 137-142; CK2_PHOSPHO_SITE 214-217; MYRISTYL 89-94;	sp_Q9BTB6_Q9BTB6_HUMAN 158-262; aminotran_4 1-291; AA_TRANSFER_CLASS_4 178-212;
DEX0449_020.orf.2	N	0 - o1-485;	224-247,1.145; 328-336,1.146; 145-152,1.082; 339-358,1.151; 120-131,1.063; 207-213,1.078; 371-397,1.213; 272-318,1.255; 361-368,1.099; 176-195,1.149; 257-268,1.195; 404-412,1.107; 464-476,1.259; 104-118,1.181; 427-449,1.147; 84-102,1.102; 26-33,1.101; 4-18,1.182;	MYRISTYL 46-51; MYRISTYL 460-465; PKC_PHOSPHO_SITE 481-483; ASN_GLYCOSYLATION 178-181; MYRISTYL 47-52; CK2_PHOSPHO_SITE 393-396; MYRISTYL 55-60; PKC_PHOSPHO_SITE 135-137; MYRISTYL 316-321; MYRISTYL 43-48; PKC_PHOSPHO_SITE 124-126; MYRISTYL 77-82; PKC_PHOSPHO_SITE 56-58; PKC_PHOSPHO_SITE 450-452; MYRISTYL 100-105; CK2_PHOSPHO_SITE 450-453; CK2_PHOSPHO_SITE 368-371; MYRISTYL 268-273; ASN_GLYCOSYLATION 33-36; MYRISTYL 76-81; CK2_PHOSPHO_SITE 153-156;	aminotran_4 170-453; sp_Q9BTB6_Q9BTB6_HUMAN 337-401; AA_TRANSFER_CLASS_4 357-391;

DEX0449_020.aa.2	N	0 - o1-465;	11-25,1.181; 394-400,1.09; 451-457,1.042; 52-59,1.082; 164-175,1.195; 319-324,1.061; 371-389,1.133; 361-368,1.14; 246-265,1.151; 278-304,1.213; 114-120,1.078; 4-9,1.069; 179-225,1.255; 235-243,1.146; 411-427,1.253; 440-446,1.08; 83-102,1.149; 27-38,1.063; 334-357,1.22; 268-275,1.099; 131-154,1.145;	ASN_GLYCOSYLATION 85-88; MYRISTYL 7-12; TYR_PHOSPHO_SITE 447-454; CK2_PHOSPHO_SITE 314-317; MYRISTYL 175-180; CK2_PHOSPHO_SITE 275-278; CK2_PHOSPHO_SITE 60-63; CK2_PHOSPHO_SITE 300-303; PKC_PHOSPHO_SITE 31-33; MYRISTYL 333-338; MYRISTYL 323-328; MYRISTYL 410-415; PKC_PHOSPHO_SITE 42-44; MYRISTYL 223-228; PKC_PHOSPHO_SITE 392-394; CK2_PHOSPHO_SITE 307-310;	sp_Q9BTB6_Q9BTB6_HUMAN 382-421; sp_Q9BTB6_Q9BTB6_HUMAN 244-310; aminotran_4 77-450; AA_TRANSFER_CLASS_4 264-298;
DEX0449_021.orf.1	N	0 - o1-74;		ASN_GLYCOSYLATION 7-10; MYRISTYL 8-13; PKC_PHOSPHO_SITE 11-13; PKC_PHOSPHO_SITE 13-15; CK2_PHOSPHO_SITE 1-4; CAMP_PHOSPHO_SITE 17-20; MYRISTYL 32-37; ASN_GLYCOSYLATION 10-13; PKC_PHOSPHO_SITE 39-41; PKC_PHOSPHO_SITE 1-3; CK2_PHOSPHO_SITE 11-14; ASN GLYCOSYLATION	

				9-12; MYRISTYL 3-8; PKC_PHOSPHO_SITE 12-14;	
DEX0449_021.aa.1	N	1 - i1-12;tm13-32;o33-63;	50-60,1.136; 4-47,1.195;		
DEX0449_022.aa.1	N	0 - o1-313;	119-137,1.178; 217-247,1.231; 76-83,1.102; 24-30,1.066; 267-282,1.131; 88-116,1.315; 250-257,1.128; 294-310,1.104; 149-155,1.04; 165-210,1.137; 139-147,1.071; 65-71,1.081; 40-59,1.204;	PKC_PHOSPHO_SITE 12-14; MYRISTYL 246-251; CK2_PHOSPHO_SITE 7-10; MYRISTYL 65-70; PKC_PHOSPHO_SITE 7-9; CK2_PHOSPHO_SITE 143-146; PKC_PHOSPHO_SITE 118-120; PKC_PHOSPHO_SITE 269-271; CK2_PHOSPHO_SITE 15-18; ASN_GLYCOSYLATION 22-25; CK2_PHOSPHO_SITE 255-258; CK2_PHOSPHO_SITE 159-162; PKC_PHOSPHO_SITE 147-149; CK2_PHOSPHO_SITE 113-116; PKC_PHOSPHO_SITE 193-195; CK2_PHOSPHO_SITE 147-150; ASN_GLYCOSYLATION 111-114; CK2_PHOSPHO_SITE 269-272; PKC_PHOSPHO_SITE 88-90; MYRISTYL 289-294; PKC_PHOSPHO_SITE 292-294; CK2_PHOSPHO_SITE 61-64;	Fe_hyd_lg_C 79-299;
DEX0449_022.orf.2	N	0 - o1-395;	345-392,1.137; 87-118,1.108; 245-251,1.081; 256-263,1.102; 220-	CK2_PHOSPHO_SITE 323-326; PKC_PHOSPHO_SITE 78-80; PKC_PHOSPHO_SITE 298-300; PKC_PHOSPHO_SITE 46-48; MYRISTYL 110-115; MYRISTYL	Fe_hyd_lg_C 259-395; CLOACIN 159-180; CLOACIN 270-289; PRO_RICH 18-105;

			239,1.204; 122-143,1.17; 319-327,1.071; 268-296,1.315; 39-58,1.105; 26-33,1.048; 171-186,1.11; 204-210,1.066; 157-169,1.174; 60-65,1.04; 299-317,1.178; 329-335,1.04;	45-50; MYRISTYL 245-250; CK2_PHOSPHO_SITE 8-11; CAMP_PHOSPHO_SITE 5-8; ASN_GLYCOSYLATION 291-294; CK2_PHOSPHO_SITE 241-244; CK2_PHOSPHO_SITE 146-149; CAMP_PHOSPHO_SITE 80-83; CK2_PHOSPHO_SITE 83-86; CK2_PHOSPHO_SITE 195-198; CK2_PHOSPHO_SITE 187-190; CK2_PHOSPHO_SITE 35-38; PKC_PHOSPHO_SITE 192-194; PKC_PHOSPHO_SITE 268-270; CK2_PHOSPHO_SITE 293-296; MYRISTYL 141-146; ASN_GLYCOSYLATION 202-205; CK2_PHOSPHO_SITE 339-342; PKC_PHOSPHO_SITE 373-375; MYRISTYL 159-164; MYRISTYL 173-178; CK2_PHOSPHO_SITE 327-330; PKC_PHOSPHO_SITE 327-329; PKC_PHOSPHO_SITE 187-189;	-
DEX0449_022.aa.2	N	0 - o1-285;	233-241,1.12; 193-207,1.205; 125-134,1.092; 166-176,1.158; 184-189,1.045; 46-76,1.231; 4-10,1.133; 145-150,1.073; 79-86,1.128; 222-	CK2_PHOSPHO_SITE 84-87; TYR_PHOSPHO_SITE 227-233; PKC_PHOSPHO_SITE 260-262; MYRISTYL 31-36; PKC_PHOSPHO_SITE 148-150; MYRISTYL 75-80; PKC_PHOSPHO_SITE 98-100; MYRISTYL 205-210; MYRISTYL 10-15; CK2_PHOSPHO_SITE 35-38:	Fe_hyd_lg_C 1-214; Fe_hyd_SSU 220-275;

			227,1.076; 96-110,1.092; 244-253,1.118; 38-44,1.11; 263-271,1.105;	CK2_PHOSPHO_SITE 117-120; MYRISTYL 206-211; CK2_PHOSPHO_SITE 98-101;	
DEX0449_023.orf.1	N	0 - o1-170;	133-143,1.109; 50-73,1.14; 81-115,1.154; 36-48,1.177; 12-30,1.165; 148-153,1.079;	PKC_PHOSPHO_SITE 161-163; MYRISTYL 143-148; PKC_PHOSPHO_SITE 118-120; PKC_PHOSPHO_SITE 119-121; MYRISTYL 127-132; CAMP_PHOSPHO_SITE 129-132; PKC_PHOSPHO_SITE 128-130; PKC_PHOSPHO_SITE 50-52;	
DEX0449_023.aa.1	N	0 - i1-159;	25-37,1.177; 122-132,1.109; 39-62,1.14; 137-142,1.079; 70-104,1.154; 4-19,1.165;	PKC_PHOSPHO_SITE 117-119; PKC_PHOSPHO_SITE 107-109; PKC_PHOSPHO_SITE 150-152; MYRISTYL 132-137; CAMP_PHOSPHO_SITE 118-121; PKC_PHOSPHO_SITE 39-41; PKC_PHOSPHO_SITE 108-110; MYRISTYL 116-121;	
DEX0449_023.orf.2	N	0 - o1-170;	81-115,1.154; 133-143,1.109; 12-30,1.165; 148-153,1.079; 50-73,1.14; 36-48,1.177;	PKC_PHOSPHO_SITE 161-163; PKC_PHOSPHO_SITE 50-52; MYRISTYL 143-148; PKC_PHOSPHO_SITE 119-121; CAMP_PHOSPHO_SITE 129-132; PKC_PHOSPHO_SITE 128-130; PKC_PHOSPHO_SITE 118-120; MYRISTYL 127-132;	
DEX0449_024.orf.1	N	0 - o1-255;	136-145,1.106; 228-249,1.154; 182-192,1.188; 4-20,1.125; 204-	PKC_PHOSPHO_SITE 150-152; MYRISTYL 228-233; MYRISTYL 225-230; CK2_PHOSPHO_SITE 241-244; CK2_PHOSPHO_SITE 250-253;	

			210,1.043; 213-224,1.184; 84-90,1.092; 28-78,1.171; 117-123,1.077; 106-115,1.076;	PKC_PHOSPHO_SITE 24-26; MYRISTYL 234-239; PKC_PHOSPHO_SITE 92-94; MYRISTYL 170-175; PKC_PHOSPHO_SITE 81-83; MYRISTYL 55-60; CK2_PHOSPHO_SITE 200-203; PKC_PHOSPHO_SITE 166-168; CK2_PHOSPHO_SITE 92-95; CK2_PHOSPHO_SITE 24-27; MYRISTYL 210-215; MYRISTYL 230-235; ASN_GLYCOSYLATION 98-101;	
DEX0449_024.aa.1	N	0 - o1-147;	35-44,1.174; 109-115,1.09; 97-105,1.083; 121-134,1.19; 5-23,1.121; 87-94,1.123; 52-74,1.216;	MYRISTYL 65-70; CK2_PHOSPHO_SITE 93-96; PKC_PHOSPHO_SITE 54-56; CK2_PHOSPHO_SITE 80-83; PKC_PHOSPHO_SITE 23-25; CK2_PHOSPHO_SITE 135-138; PKC_PHOSPHO_SITE 108-110; PKC_PHOSPHO_SITE 45-47;	CSD 62-129; CSP 64-129; COLDSHOCK 65-80; COLDSHOCK 86-95;
DEX0449_024.orf.2	N	0 - o1-255;	213-224,1.184; 182-192,1.188; 84-90,1.092; 117-123,1.077; 28-78,1.171; 136-145,1.106; 4-20,1.125; 204-210,1.043; 228-249,1.154; 106-115,1.076;	CK2_PHOSPHO_SITE 200-203; PKC_PHOSPHO_SITE 24-26; CK2_PHOSPHO_SITE 241-244; PKC_PHOSPHO_SITE 150-152; CK2_PHOSPHO_SITE 24-27; MYRISTYL 225-230; MYRISTYL 230-235; MYRISTYL 55-60; MYRISTYL 234-239; MYRISTYL 170-175; PKC_PHOSPHO_SITE 92-94; PKC_PHOSPHO_SITE 166-168; CK2_PHOSPHO_SITE 250-253; MYRISTYL 228-233; ASN GLYCOSYLATION	

				98-101; PKC_PHOSPHO_SITE 81-83; CK2_PHOSPHO_SITE 92-95; MYRISTYL 210-215;	
DEX0449_024.orf.3	N	0 - o1-243;	94-103,1.076; 192-198,1.043; 170-180,1.188; 216-237,1.154; 16-65,1.171; 72-78,1.092; 201-212,1.184; 124-133,1.106; 105-111,1.077;	MYRISTYL 9-14; CK2_PHOSPHO_SITE 238-241; CK2_PHOSPHO_SITE 229-232; MYRISTYL 216-221; CK2_PHOSPHO_SITE 188-191; MYRISTYL 222-227; MYRISTYL 218-223; MYRISTYL 8-13; PKC_PHOSPHO_SITE 69-71; MYRISTYL 213-218; PKC_PHOSPHO_SITE 154-156; MYRISTYL 43-48; PKC_PHOSPHO_SITE 138-140; MYRISTYL 158-163; PKC_PHOSPHO_SITE 80-82; ASN_GLYCOSYLATION 86-89; CK2_PHOSPHO_SITE 80-83; PKC_PHOSPHO_SITE 3-5; MYRISTYL 198-203;	
DEX0449_024.aa.3	N	0 - o1-147;	5-23,1.121; 52-74,1.216; 121-134,1.19; 109-115,1.09; 35-44,1.174; 97-105,1.083; 87-94,1.123;	PKC_PHOSPHO_SITE 54-56; CK2_PHOSPHO_SITE 80-83; PKC_PHOSPHO_SITE 45-47; CK2_PHOSPHO_SITE 135-138; PKC_PHOSPHO_SITE 23-25; CK2_PHOSPHO_SITE 93-96; MYRISTYL 65-70; PKC_PHOSPHO_SITE 108-110;	CSP 64-129; COLDSHOCK 65-80; COLDSHOCK 86-95; CSD 62-129;
DEX0449_024.orf.4	N	0 - o1-202;	129-139,1.188; 160-171,1.184; 23-37,1.092; 151-157,1.043; 53-62,1.076; 175-	MYRISTYL 181-186; MYRISTYL 175-180; ASN_GLYCOSYLATION 45-48; PKC_PHOSPHO_SITE 29-31; MYRISTYL 172-177; CK2_PHOSPHO_SITE 188-191;	

			196,1.154; 64-70,1.077; 11-17,1.151; 83-92,1.106;	PKC_PHOSPHO_SITE 39-41; MYRISTYL 117-122; MYRISTYL 2-7; PKC_PHOSPHO_SITE 113-115; CK2_PHOSPHO_SITE 39-42; CK2_PHOSPHO_SITE 197-200; CK2_PHOSPHO_SITE 147-150; MYRISTYL 157-162; PKC_PHOSPHO_SITE 97-99; MYRISTYL 177-182;	
DEX0449_024.aa.4	N	0 - o1-138;	108-119,1.184; 99-105,1.043; 123-135,1.126; 4-10,1.076; 12-18,1.077; 31-40,1.106; 77-87,1.188;	CK2_PHOSPHO_SITE 95-98; MYRISTYL 105-110; PKC_PHOSPHO_SITE 45-47; MYRISTYL 123-128; MYRISTYL 120-125; PKC_PHOSPHO_SITE 61-63; MYRISTYL 65-70;	
DEX0449_024.aa.5	N	0 - o1-185;	124-134,1.188; 78-87,1.106; 155-166,1.184; 59-65,1.077; 48-57,1.076; 170-182,1.126; 146-152,1.043; 5-23,1.121;	CK2_PHOSPHO_SITE 142-145; MYRISTYL 170-175; CK2_PHOSPHO_SITE 34-37; PKC_PHOSPHO_SITE 92-94; PKC_PHOSPHO_SITE 34-36; MYRISTYL 152-157; PKC_PHOSPHO_SITE 23-25; PKC_PHOSPHO_SITE 108-110; MYRISTYL 112-117; MYRISTYL 167-172; ASN_GLYCOSYLATION 40-43;	
DEX0449_024.orf.5	N	0 - o1-265;	214-220,1.043; 4-14,1.143; 223-234,1.184; 146-155,1.106; 127-133,1.077; 116-125,1.076; 73-91,1.121; 192-202,1.188; 57-65,1.074;	PKC_PHOSPHO_SITE 102-104; PKC_PHOSPHO_SITE 176-178; MYRISTYL 27-32; PKC_PHOSPHO_SITE 91-93; MYRISTYL 235-240; MYRISTYL 31-36; PKC_PHOSPHO_SITE 51-53; MYRISTYL 238-243; PKC_PHOSPHO_SITE 160-162; CK2_PHOSPHO_SITE	

			238-259,1.154; 18-27,1.167;	260-263; MYRISTYL 180-185; MYRISTYL 240-245; CK2_PHOSPHO_SITE 102-105; MYRISTYL 42-47; CK2_PHOSPHO_SITE 210-213; MYRISTYL 220-225; CK2_PHOSPHO_SITE 251-254; MYRISTYL 244-249; PKC_PHOSPHO_SITE 15-17; ASN_GLYCOSYLATION 108-111;	
DEX0449_024.aa.6	Y	0 - o1-201;	42-56,1.099; 33-39,1.074; 58-181,1.229; 14-31,1.197;	MYRISTYL 95-100; PKC_PHOSPHO_SITE 30-32; CK2_PHOSPHO_SITE 127-130; MYRISTYL 143-148; CK2_PHOSPHO_SITE 10-13; PKC_PHOSPHO_SITE 163-165; PKC_PHOSPHO_SITE 10-12; CK2_PHOSPHO_SITE 74-77; MYRISTYL 67-72; PKC_PHOSPHO_SITE 181-183;	ER_TARGET 198-201;
DEX0449_025.aa.1	N	0 - o1-159;	149-155,1.129; 17-32,1.132; 109-136,1.177; 74-88,1.162; 49-67,1.112; 99-107,1.157; 34-43,1.178;	CK2_PHOSPHO_SITE 138-141; CK2_PHOSPHO_SITE 41-44; PKC_PHOSPHO_SITE 85-87; MYRISTYL 144-149; CK2_PHOSPHO_SITE 91-94; PKC_PHOSPHO_SITE 52-54; CK2_PHOSPHO_SITE 108-111; PKC_PHOSPHO_SITE 29-31; CK2_PHOSPHO_SITE 143-146;	
DEX0449_025.aa.2	N	1 - i1-36;tm37-54;o55-128;	35-44,1.203; 4-10,1.103; 118-124,1.129; 48-56,1.102; 68-76,1.157; 78-105,1.177;	CK2_PHOSPHO_SITE 77-80; CK2_PHOSPHO_SITE 60-63; ASN_GLYCOSYLATION 44-47; MYRISTYL 113-118; CK2_PHOSPHO_SITE 107-110:	

				PKC_PHOSPHO_SITE 57-59; ASN_GLYCOSYLATION 12-15; CK2_PHOSPHO_SITE 112-115; PKC_PHOSPHO_SITE 13-15;	
DEX0449_026.aa.1	N	0 - o1-224;	63-74,1.101; 148-166,1.202; 43-49,1.048; 186-195,1.118; 135-141,1.07; 16-28,1.137; 124-129,1.03; 114-120,1.096; 170-178,1.135; 198-205,1.103; 90-96,1.08; 212-221,1.109;	CK2_PHOSPHO_SITE 119-122; CAMP_PHOSPHO_SITE 143-146; CK2_PHOSPHO_SITE 188-191; ASN_GLYCOSYLATION 59-62; PKC_PHOSPHO_SITE 77-79; CK2_PHOSPHO_SITE 101-104; PKC_PHOSPHO_SITE 141-143; CK2_PHOSPHO_SITE 108-111; PKC_PHOSPHO_SITE 37-39; TYR_PHOSPHO_SITE 166-174; PKC_PHOSPHO_SITE 142-144;	SH2DOMAIN 113-127; SH2 113-196; SH2 111-202; PRO_RICH 19-102; SH2 113-221; sp_Q13094_LCP2_HUMAN 113-207; SH2DOMAIN 185-199; SH2DOMAIN 133-143;
DEX0449_026.aa.2	Y	0 - o1-230;	176-184,1.135; 154-172,1.202; 120-126,1.096; 49-55,1.048; 218-227,1.109; 204-211,1.103; 141-147,1.07; 69-80,1.101; 8-34,1.204; 130-135,1.03; 192-201,1.118; 96-102,1.08;	TYR_PHOSPHO_SITE 172-180; PKC_PHOSPHO_SITE 148-150; PKC_PHOSPHO_SITE 43-45; CK2_PHOSPHO_SITE 114-117; PKC_PHOSPHO_SITE 83-85; MYRISTYL 8-13; PKC_PHOSPHO_SITE 147-149; CAMP_PHOSPHO_SITE 149-152; CK2_PHOSPHO_SITE 107-110; CK2_PHOSPHO_SITE 194-197; CK2_PHOSPHO_SITE 125-128; ASN_GLYCOSYLATION 65-68;	SH2DOMAIN 119-133; PRO_RICH 25-108; sp_Q13094_LCP2_HUMAN 119-213; SH2 119-227; SH2DOMAIN 191-205; SH2 117-208; SH2 119-202; SH2DOMAIN 139-149;
DEX0449_026.orf.2	Y	0 - o1-225;	91-97,1.08; 136-142,1.07; 149-167,1.202; 213-	PKC_PHOSPHO_SITE 78-80; PKC_PHOSPHO_SITE 38-40; CK2_PHOSPHO_SITE 189-192;	SH2 114-197; SH2 114-222; sp_Q13094_LCP2_HUMAN 114-208; PRO_RICH 20-103; SH2 112-203; SH2DOMAIN 186-

			222,1.109; 187-196,1.118; 64-75,1.101; 125-130,1.03; 171-179,1.135; 199-206,1.103; 44-50,1.048; 4-29,1.249; 115-121,1.096;	CK2_PHOSPHO_SITE 102-105; PKC_PHOSPHO_SITE 142-144; CK2_PHOSPHO_SITE 120-123; ASN_GLYCOSYLATION 60-63; PKC_PHOSPHO_SITE 143-145; CK2_PHOSPHO_SITE 109-112; CAMP_PHOSPHO_SITE 144-147; TYR_PHOSPHO_SITE 167-175;	200; SH2DOMAIN 114-128; SH2DOMAIN 134-144;
DEX0449_027.aa.1	N	0 - o1-1085;	1028-1040,1.222; 860-866,1.065; 684-694,1.255; 13-34,1.134; 660-676,1.101; 1008-1020,1.107; 176-185,1.108; 784-811,1.127; 720-736,1.173; 205-223,1.113; 981-994,1.16; 36-43,1.066; 426-450,1.214; 613-624,1.087; 957-973,1.166; 585-592,1.129; 311-317,1.111; 357-369,1.131; 286-309,1.102; 704-713,1.132; 117-129,1.172; 884-894,1.151; 747-	CK2_PHOSPHO_SITE 80-83; MYRISTYL 99-104; MYRISTYL 98-103; CK2_PHOSPHO_SITE 181-184; MYRISTYL 335-340; MYRISTYL 95-100; MYRISTYL 512-517; MYRISTYL 96-101; MYRISTYL 683-688; CK2_PHOSPHO_SITE 116-119; MYRISTYL 42-47; CK2_PHOSPHO_SITE 580-583; MYRISTYL 889-894; ASN_GLYCOSYLATION 766-769; TYR_PHOSPHO_SITE 323-330; MYRISTYL 656-661; CK2_PHOSPHO_SITE 299-302; PKC_PHOSPHO_SITE 390-392; MYRISTYL 679-684; MYRISTYL 12-17; MYRISTYL 738-743; CAMP_PHOSPHO_SITE 236-239; ASN_GLYCOSYLATION 456-459; PKC_PHOSPHO_SITE 680-682; MYRISTYL 63-68; MYRISTYL 76-81; ASN_GLYCOSYLATION 783-786; PKC_PHOSPHO_SITE 939-941; MYRISTYL 93-98; PKC PHOSPHO SITE	LON 123-368; LON_SER 978-986; ENDOLAPTASE 1005-1024; GLY_RICH 56-106; LON 123-368; AAA 515-786; ENDOLAPTASE 1028-1046; ATP_GTP_A 523-530; lon 125-1073; ENDOLAPTASE 975-994; DISEASERSIST 579-593; DISEASERSIST 518-533; ENDOLAPTASE 523-542; ENDOLAPTASE 888-904; AAA 518-839;

			766,1.195; 251-257,1.062; 639-648,1.073; 516-530,1.143; 131-144,1.103; 602-608,1.102; 264-283,1.144; 815-833,1.167; 738-745,1.081; 1056-1072,1.154; 941-954,1.059; 925-934,1.076; 321-333,1.112; 480-489,1.102; 869-880,1.117; 146-169,1.172; 533-539,1.047; 499-510,1.138; 1046-1054,1.093; 571-579,1.12; 462-468,1.049; 45-53,1.059; 835-858,1.106; 563-569,1.04; 372-378,1.057; 903-909,1.107; 106-112,1.104; 392-406,1.113; 5-11,1.071;	1012-1014; MYRISTYL 634-639; PKC_PHOSPHO_SITE 975-977; MYRISTYL 89-94; CK2_PHOSPHO_SITE 975-978; AMIDATION 241-244; PKC_PHOSPHO_SITE 72-74; CK2_PHOSPHO_SITE 339-342; ASN_GLYCOSYLATION 450-453; PKC_PHOSPHO_SITE 299-301; CK2_PHOSPHO_SITE 770-773; MYRISTYL 893-898; MYRISTYL 675-680; PKC_PHOSPHO_SITE 251-253; MYRISTYL 734-739; MYRISTYL 744-749; MYRISTYL 735-740; PKC_PHOSPHO_SITE 239-241; MYRISTYL 654-659; MYRISTYL 77-82; MYRISTYL 81-86; MYRISTYL 809-814; MYRISTYL 526-531; ASN_GLYCOSYLATION 174-177; PKC_PHOSPHO_SITE 909-911; MYRISTYL 446-451; MYRISTYL 637-642; MYRISTYL 567-572;	
DEX0449_027.orf.1	N	0 - o1-668;	444-458,1.113; 615-621.1.04:	CK2_PHOSPHO_SITE 132-135; MYRISTYL 387-392; MYRISTYL 145-150:	GLY_RICH 108-153; DISEASERSIST 570-585; DISEASERSIST 631-645: ATP GTP A

			41-48,1.124; 183-196,1.103; 532-541,1.102; 585-591,1.047; 478-502,1.214; 623-631,1.12; 363-369,1.111; 424-430,1.057; 88-95,1.066; 257-275,1.113; 169-181,1.172; 551-562,1.138; 198-221,1.172; 10-16,1.06; 57-63,1.071; 514-520,1.049; 316-335,1.144; 65-86,1.134; 158-164,1.104; 30-36,1.085; 228-237,1.108; 637-644,1.129; 654-660,1.102; 409-421,1.131; 303-309,1.062; 338-361,1.102; 568-582,1.143; 373-385,1.112; 97-105,1.059;	AMIDATION 293-296; ASN_GLYCOSYLATION 226-229; MYRISTYL 39-44; MYRISTYL 141-146; PKC_PHOSPHO_SITE 303-305; CK2_PHOSPHO_SITE 632-635; CK2_PHOSPHO_SITE 233-236; MYRISTYL 133-138; CK2_PHOSPHO_SITE 391-394; CK2_PHOSPHO_SITE 168-171; MYRISTYL 128-133; MYRISTYL 147-152; MYRISTYL 150-155; PKC_PHOSPHO_SITE 291-293; CK2_PHOSPHO_SITE 23-26; MYRISTYL 148-153; PKC_PHOSPHO_SITE 12-14; CK2_PHOSPHO_SITE 38-41; MYRISTYL 64-69; PKC_PHOSPHO_SITE 124-126; MYRISTYL 578-583; ASN_GLYCOSYLATION 508-511; PKC_PHOSPHO_SITE 442-444; MYRISTYL 115-120; CAMP_PHOSPHO_SITE 288-291; MYRISTYL 564-569; MYRISTYL 498-503; MYRISTYL 619-624; CK2_PHOSPHO_SITE 351-354; PKC_PHOSPHO_SITE 351-353; CAMP_PHOSPHO_SITE 8-11; MYRISTYL 151-156; MYRISTYL 129-134; TYR_PHOSPHO_SITE 375-382; ASN_GLYCOSYLATION 502-505; PKC_PHOSPHO_SITE 6-8; MYRISTYL 94-99;	575-582; AAA 567-665; LON 175-420; LON 175-420;
DEX0449 02	N	0 - o1-	345-	TYR PHOSPHO SITE	lon 3-759: AAA 330-

7.aa.2		771;	351,1.047; 470-497,1.127; 238-262,1.214; 204-218,1.113; 611-620,1.076; 501-519,1.167; 133-145,1.112; 570-580,1.151; 643-659,1.166; 714-726,1.222; 667-680,1.16; 627-640,1.059; 521-544,1.106; 742-758,1.154; 589-595,1.107; 79-91,1.062; 732-740,1.093; 6-19,1.089; 169-181,1.131; 123-129,1.111; 184-190,1.057; 397-404,1.129; 311-322,1.138; 32-57,1.174; 432-452,1.18; 546-552,1.065; 328-342,1.143; 416-425,1.132; 95-121,1.152; 694-706,1.107; 292-301,1.102; 274-	135-142; ASN_GLYCOSYLATION 469-472; MYRISTYL 64-69; MYRISTYL 579-584; MYRISTYL 575-580; CK2_PHOSPHO_SITE 661-664; PKC_PHOSPHO_SITE 698-700; ASN_GLYCOSYLATION 452-455; CK2_PHOSPHO_SITE 151-154; ASN_GLYCOSYLATION 268-271; ASN_GLYCOSYLATION 262-265; CK2_PHOSPHO_SITE 111-114; CK2_PHOSPHO_SITE 456-459; MYRISTYL 258-263; PKC_PHOSPHO_SITE 661-663; PKC_PHOSPHO_SITE 111-113; MYRISTYL 68-73; PKC_PHOSPHO_SITE 202-204; PKC_PHOSPHO_SITE 595-597; CK2_PHOSPHO_SITE 392-395; MYRISTYL 102-107; MYRISTYL 379-384; PKC_PHOSPHO_SITE 625-627; MYRISTYL 338-343; MYRISTYL 324-329; MYRISTYL 147-152; CAMP_PHOSPHO_SITE 4-7; MYRISTYL 73-78; MYRISTYL 495-500;	525; ENDOLAPTASE 574-590; AAA 327-472; LON_SER 664-672; DISEASERSIST 391-405; LON 2-180; DISEASERSIST 330-345; ENDOLAPTASE 335-354; ENDOLAPTASE 714-732; ENDOLAPTASE 691-710; LON 5-180; CLPPROTEASEA 441-455; ATP_GTP_A 335-342; CLPPROTEASEA 331-349; ENDOLAPTASE 661-680;

			280,1.049; 555-566,1.117; 383-391,1.12; 375-381,1.04;		
DEX0449_027.aa.3	N	0 - o1-648;	321-333,1.112; 205-223,1.113; 499-510,1.138; 131-144,1.103; 36-43,1.066; 286-309,1.102; 533-539,1.047; 585-592,1.129; 146-169,1.172; 426-450,1.214; 117-129,1.172; 813-819,1.153; 176-185,1.108; 480-489,1.102; 658-685,1.127; 839-845,1.126; 743-754,1.117; 251-257,1.062; 604-613,1.132; 734-740,1.065; 392-406,1.113; 784-790,1.079; 13-34,1.134; 516-530,1.143; 264-283,1.144; 758-768,1.151; 709-732.1.106:	CK2_PHOSPHO_SITE 181-184; ASN_GLYCOSYLATION 174-177; MYRISTYL 810-815; MYRISTYL 526-531; MYRISTYL 93-98; MYRISTYL 767-772; MYRISTYL 76-81; CK2_PHOSPHO_SITE 116-119; ASN_GLYCOSYLATION 657-660; MYRISTYL 512-517; AMIDATION 241-244; CK2_PHOSPHO_SITE 80-83; ASN_GLYCOSYLATION 640-643; MYRISTYL 89-94; CAMP_PHOSPHO_SITE 236-239; MYRISTYL 446-451; MYRISTYL 763-768; MYRISTYL 98-103; MYRISTYL 567-572; PKC_PHOSPHO_SITE 239-241; MYRISTYL 824-829; CK2_PHOSPHO_SITE 299-302; MYRISTYL 99-104; MYRISTYL 12-17; MYRISTYL 683-688; MYRISTYL 42-47; MYRISTYL 95-100; ASN_GLYCOSYLATION 450-453; MYRISTYL 77-82; PKC_PHOSPHO_SITE 72-74; MYRISTYL 81-86; MYRISTYL 96-101; PKC_PHOSPHO_SITE 299-301; MYRISTYL 63-68; TYR_PHOSPHO_SITE 323-330; PKC_PHOSPHO_SITE 251-253; PKC PHOSPHO SITE	lon 125-842; AAA 515-660; DISEASERSIST 579-593; LON 123-368; CLPPROTEASEA 629-643; LIPOCALIN 803-815; AAA 518-713; ENDOLAPTASE 523-542; ENDOLAPTASE 762-778; CLPPROTEASEA 519-537; ATP_GTP_A 523-530; GLY_RICH 56-106; DISEASERSIST 518-533; LON 123-368;

			563-569,1.04; 357-369,1.131; 462-468,1.049; 311-317,1.111; 571-579,1.12; 45-53,1.059; 794-809,1.217; 620-640,1.18; 372-378,1.057; 106-112,1.104; 689-707,1.167; 5-11,1.071;	792-794; AMIDATION 778-781; CK2_PHOSPHO_SITE 580-583; CK2_PHOSPHO_SITE 339-342; ASN_GLYCOSYLATION 456-459; CK2_PHOSPHO_SITE 644-647; MYRISTYL 335-340; PKC_PHOSPHO_SITE 390-392;	
DEX0449_028.aa.1	N	0 - o1-124;	4-14,1.157; 110-116,1.024; 37-45,1.061; 48-53,1.048; 62-79,1.175;	MYRISTYL 115-120; CK2_PHOSPHO_SITE 15-18; TYR_PHOSPHO_SITE 58-66; CK2_PHOSPHO_SITE 99-102; PKC_PHOSPHO_SITE 37-39; TYR_PHOSPHO_SITE 45-53; ASN_GLYCOSYLATION 90-93; CK2_PHOSPHO_SITE 30-33;	
DEX0449_029.aa.1	N	1 - i1-16;tm17-34;o35-296;	11-36,1.154; 181-194,1.132; 261-267,1.144; 213-221,1.069; 95-119,1.183; 60-70,1.115; 247-254,1.114; 148-161,1.085; 125-139,1.129; 167-177,1.148;	CK2_PHOSPHO_SITE 242-245; PKC_PHOSPHO_SITE 3-5; MYRISTYL 258-263; CK2_PHOSPHO_SITE 40-43; PKC_PHOSPHO_SITE 219-221; PKC_PHOSPHO_SITE 146-148; MYRISTYL 155-160; ASN_GLYCOSYLATION 239-242; PKC_PHOSPHO_SITE 179-181; CK2_PHOSPHO_SITE 195-198; MYRISTYL 22-27; CK2_PHOSPHO_SITE 179-182; CAMP_PHOSPHO_SITE	KH 51-120; KH_TYPE_1_1 52-115; KH 56-103; KH 128-178; KH 123-195; KH_TYPE_1_2 124-190;

				79-82;	
DEX0449_030.aa.1	N	1 - i1-164;tm165-187;o188-195;	11-17,1.104; 83-101,1.178; 140-151,1.186; 24-32,1.128; 42-78,1.18; 114-126,1.137; 165-192,1.253; 105-112,1.023;	PKC_PHOSPHO_SITE 161-163; PKC_PHOSPHO_SITE 99-101; MYRISTYL 181-186; ASN_GLYCOSYLATION 21-24; PKC_PHOSPHO_SITE 68-70; MYRISTYL 185-190; MYRISTYL 158-163; AMIDATION 191-194; ASN_GLYCOSYLATION 129-132; MYRISTYL 171-176; CK2_PHOSPHO_SITE 131-134; ASN_GLYCOSYLATION 93-96;	MHCCLASSI 136-154; MHCCLASSI 65-81;
DEX0449_030.orf.1	N	1 - o1-165;tm166-188;i189-194;		ASN_GLYCOSYLATION 128-131; AMIDATION 190-193; MYRISTYL 170-175; CK2_PHOSPHO_SITE 3-6; PKC_PHOSPHO_SITE 67-69; AMIDATION 3-6; ASN_GLYCOSYLATION 92-95; MYRISTYL 157-162; CK2_PHOSPHO_SITE 130-133; MYRISTYL 184-189; PKC_PHOSPHO_SITE 3-5; PKC_PHOSPHO_SITE 98-100; ASN_GLYCOSYLATION 20-23; MYRISTYL 7-12; MYRISTYL 180-185; PKC_PHOSPHO_SITE 160-162; CK2_PHOSPHO_SITE 7-10; ASN_GLYCOSYLATION 6-9; CAMP_PHOSPHO_SITE 5-8; MYRISTYL 2-7;	MHCCLASSI 64-80; MHCCLASSI 135-153;
DEX0449_031.aa.1	N	0 - o1-132;	122-129,1.15; 10-24,1.133; 67-74,1.074; 105-	MYRISTYL 22-27; PKC_PHOSPHO_SITE 33-35; ASN_GLYCOSYLATION 17-20;	

			117,1.25; 94-102,1.038;	ASN_GLYCOSYLATION 2-5; PKC_PHOSPHO_SITE 118-120; MYRISTYL 61-66; PKC_PHOSPHO_SITE 4-6; CK2_PHOSPHO_SITE 4-7; AMIDATION 103-106; MYRISTYL 89-94; CAMP_PHOSPHO_SITE 76-79; PKC_PHOSPHO_SITE 32-34; CK2_PHOSPHO_SITE 88-91; ASN_GLYCOSYLATION 27-30;	
DEX0449_031.orf.1	N	0 - o1-199;	141-146,1.033; 46-52,1.027; 54-61,1.105; 153-163,1.116; 171-192,1.152; 94-101,1.084;	CK2_PHOSPHO_SITE 192-195; PKC_PHOSPHO_SITE 166-168; PKC_PHOSPHO_SITE 3-5; PKC_PHOSPHO_SITE 148-150; PKC_PHOSPHO_SITE 102-104; PKC_PHOSPHO_SITE 20-22; PKC_PHOSPHO_SITE 184-186; PKC_PHOSPHO_SITE 50-52; CK2_PHOSPHO_SITE 148-151; ASN_GLYCOSYLATION 167-170; CK2_PHOSPHO_SITE 79-82; CAMP_PHOSPHO_SITE 12-15; ASN_GLYCOSYLATION 77-80; PKC_PHOSPHO_SITE 79-81;	LYS_RICH 5-40;
DEX0449_032.aa.1	N	0 - o1-259;	122-131,1.066; 136-174,1.19; 100-106,1.068; 82-89,1.039; 190-197,1.171; 232-238,1.086; 204-	ASN_GLYCOSYLATION 130-133; MYRISTYL 98-103; MYRISTYL 53-58; CK2_PHOSPHO_SITE 132-135; MYRISTYL 52-57; AMIDATION 2-5; PKC_PHOSPHO_SITE 250-252; PKC_PHOSPHO_SITE 45-47; MYRISTYL	UBCc 111-255; UQ_con 107-250; UBIQUITIN_CONJUGAT_1 181-195; sp_O14933_UBC8_HUMAN 111-250; UBIQUITIN_CONJUGAT_2 111-244;

			220,1.175; 177-184,1.082; 10-16,1.056; 21-53,1.139;	17-22; MYRISTYL 65-70; CK2_PHOSPHO_SITE 115-118; PKC_PHOSPHO_SITE 177-179; MYRISTYL 74-79; CK2_PHOSPHO_SITE 93-96; PKC_PHOSPHO_SITE 93-95; CK2_PHOSPHO_SITE 113-116; MYRISTYL 70-75; MYRISTYL 69-74;	
DEX0449_032.aa.2	N	0 - i1-156;	87-94,1.171; 129-135,1.086; 74-81,1.082; 101-117,1.175; 19-28,1.066; 33-71,1.19;	PKC_PHOSPHO_SITE 147-149; PKC_PHOSPHO_SITE 74-76; CAMP_PHOSPHO_SITE 2-5; ASN_GLYCOSYLATION 27-30; CK2_PHOSPHO_SITE 29-32; PKC_PHOSPHO_SITE 5-7; CK2_PHOSPHO_SITE 12-15;	UBCc 8-152; UBIQUITIN_CONJUGAT_1 78-92; UBIQUITIN_CONJUGAT_2 13-141; sp_014933_UBC8_HUMAN 13-147; UQ_con 11-147;
DEX0449_032.orf.2	N	0 - o1-168;	31-40,1.066; 45-83,1.19; 4-10,1.1; 99-106,1.171; 141-147,1.086; 86-93,1.082; 113-129,1.175;	PKC_PHOSPHO_SITE 159-161; CK2_PHOSPHO_SITE 41-44; ASN_GLYCOSYLATION 39-42; MYRISTYL 16-21; PKC_PHOSPHO_SITE 86-88;	UBIQUITIN_CONJUGAT_2 20-153; UBCc 20-164; UQ_con 15-159; sp_014933_UBC8_HUMAN 25-159; UBIQUITIN_CONJUGAT_1 90-104;
DEX0449_032.aa.3	N	0 - o1-137;	82-98,1.175; 68-75,1.171; 110-116,1.086; 14-52,1.19; 55-62,1.082;	AMIDATION 2-5; PKC_PHOSPHO_SITE 55-57; ASN_GLYCOSYLATION 8-11; PKC_PHOSPHO_SITE 128-130; MYRISTYL 7-12; CK2_PHOSPHO_SITE 10-13;	UBCc 3-133; UBIQUITIN_CONJUGAT_2 12-122; UBIQUITIN_CONJUGAT_1 59-73; sp_014933_UBC8_HUMAN 8-128; UQ_con 1-128;
DEX0449_032.aa.4	N	0 - o1-237;	155-162,1.082; 113-118,1.041; 168-175,1.171; 121-152,1.095; 75-82,1.124;	ASN_GLYCOSYLATION 24-27; PKC_PHOSPHO_SITE 65-67; MYRISTYL 85-90; PKC_PHOSPHO_SITE 155-157; PKC_PHOSPHO_SITE 4-6;	sp_014933_UBC8_HUMAN 126-228; sp_014933_UBC8_HUMAN 3-42; UQ_con 1-228; UBIQUITIN_CONJUGAT_1 159-173; UBIQUITIN_CONJUGAT_2 122-222; UBCc 5-

			95-101,1.1; 30-58,1.19; 182-198,1.175; 210-216,1.086; 16-25,1.066; 5-14,1.074;	PKC_PHOSPHO_SITE 228-230; CK2_PHOSPHO_SITE 26-29; PKC_PHOSPHO_SITE 118-120;	233;
DEX0449_032.orf.4	N	0 - o1-156;	32-37,1.041; 129-135,1.086; 101-117,1.175; 74-81,1.082; 87-94,1.171; 40-71,1.095;	CK2_PHOSPHO_SITE 16-19; PKC_PHOSPHO_SITE 147-149; PKC_PHOSPHO_SITE 74-76; PKC_PHOSPHO_SITE 37-39; MYRISTYL 14-19;	UBIQUITIN_CONJUGAT_2 41-141; UQ_con 14-147; UBIQUITIN_CONJUGAT_1 78-92; UBCc 12-152; sp_O14933_UBC8_HUMAN 45-147;
DEX0449_033.aa.1	N	0 - o1-88;	51-58,1.097; 4-12,1.097; 23-30,1.121;	PKC_PHOSPHO_SITE 38-40; CK2_PHOSPHO_SITE 48-51; ASN_GLYCOSYLATION 79-82;	
DEX0449_033.aa.2	N	0 - o1-40;	21-27,1.087; 31-37,1.071;	MYRISTYL 7-12; AMIDATION 10-13;	
DEX0449_033.orf.2	N	0 - o1-60;	21-39,1.171; 11-17,1.065;	MYRISTYL 17-22; MYRISTYL 55-60;	
DEX0449_034.aa.1	N	0 - o1-213;		PKC_PHOSPHO_SITE 77-79; CK2_PHOSPHO_SITE 64-67; MYRISTYL 108-113; PKC_PHOSPHO_SITE 209-211; CK2_PHOSPHO_SITE 161-164; CK2_PHOSPHO_SITE 122-125; MYRISTYL 23-28; CK2_PHOSPHO_SITE 113-116; MYRISTYL 5-10; CK2_PHOSPHO_SITE 77-80; PKC_PHOSPHO_SITE 64-66; MYRISTYL 73-78; CK2_PHOSPHO_SITE 209-212;	
DEX0449_035.aa.1	N	0 - o1-458;	223-245,1.186; 187-196,1.158; 175-180,1.057; 427-436,1.181; 279-	ASN_GLYCOSYLATION 328-331; MYRISTYL 243-248; PKC_PHOSPHO_SITE 352-354; ASN_GLYCOSYLATION 171-174; PKC_PHOSPHO_SITE 420-422;	ANK 352-381; ANK 219-248; ANK 388-421; ank 291-350; ANK_REPEAT_3 388-424; ank 183-215; ank 388-424; ANK 183-212; ANK 291-321; ANKYRIN 405-417; ANK_REPEAT_2

			286,1.076; 256-263,1.17; 28-61,1.087; 72-85,1.167; 366-385,1.175; 198-205,1.169; 108-114,1.091; 293-310,1.204; 316-347,1.133; 119-125,1.087; 154-162,1.094; 265-276,1.128; 355-361,1.101; 392-412,1.161; 100-106,1.066;	CK2_PHOSPHO_SITE 330-333; ASN_GLYCOSYLATION 149-152; ASN_GLYCOSYLATION 77-80; AMIDATION 194-197; ASN_GLYCOSYLATION 247-250; PKC_PHOSPHO_SITE 43-45; MYRISTYL 275-280; MYRISTYL 416-421; CK2_PHOSPHO_SITE 120-123; MYRISTYL 145-150; ASN_GLYCOSYLATION 136-139; CK2_PHOSPHO_SITE 288-291; CK2_PHOSPHO_SITE 18-21; PKC_PHOSPHO_SITE 334-336; CK2_PHOSPHO_SITE 404-407;	352-384; ank 352-384; ANK_REPEAT_1 219-251; ANK_REP_REGION 183-432; ank 252-290; ank 219-251; ANKYRIN 253-265; ANK 252-283;
DEX0449_035.orf.1	N	0 - o1-439;	274-291,1.204; 100-106,1.087; 168-177,1.158; 179-186,1.169; 297-328,1.133; 4-42,1.154; 246-257,1.128; 204-226,1.186; 336-342,1.101; 347-366,1.175; 53-66,1.167; 260-267,1.076; 135-143,1.094; 156-161,1.057; 237-244,1.17; 408-417,1.181; 81-87,1.066; 373-	ASN_GLYCOSYLATION 309-312; ASN_GLYCOSYLATION 130-133; MYRISTYL 224-229; ASN_GLYCOSYLATION 152-155; ASN_GLYCOSYLATION 228-231; MYRISTYL 256-261; CK2_PHOSPHO_SITE 101-104; AMIDATION 175-178; PKC_PHOSPHO_SITE 315-317; PKC_PHOSPHO_SITE 333-335; MYRISTYL 126-131; MYRISTYL 397-402; PKC_PHOSPHO_SITE 401-403; ASN_GLYCOSYLATION 58-61; CK2_PHOSPHO_SITE 311-314; PKC_PHOSPHO_SITE 24-26; CK2_PHOSPHO_SITE 385-388; CK2_PHOSPHO_SITE 269-272;	ANK 272-302; ANK 369-402; ANK_REPEAT_1 200-232; ANKYRIN 234-246; ANKYRIN 386-398; ANK_REPEAT_3 369-405; ANK 233-264; ANK 200-229; ank 272-331; ANK 164-193; ANK_REP_REGION 164-413; ank 200-232; ANK_REPEAT_2 333-365; ank 333-365; ank 164-196; ank 369-405; ank 233-271; ANK 333-362;

			393,1.161; 89-95,1.091;	ASN_GLYCOSYLATION 117-120;	
DEX0449_036.aa.1	Y	0 - o1-130;	4-21,1.165; 23-43,1.129; 110-127,1.174; 49-67,1.178; 82-102,1.143;	CK2_PHOSPHO_SITE 68-71; PKC_PHOSPHO_SITE 81-83; AMIDATION 74-77; MYRISTYL 19-24; CK2_PHOSPHO_SITE 81-84; CAMP_PHOSPHO_SITE 76-79; PKC_PHOSPHO_SITE 127-129;	
DEX0449_036.orf.1	N	0 - o1-243;	157-196,1.173; 83-93,1.055; 215-223,1.167; 11-37,1.163; 139-149,1.056; 225-240,1.175; 110-116,1.067; 49-61,1.064; 119-131,1.061; 201-208,1.12;	MYRISTYL 135-140; PKC_PHOSPHO_SITE 173-175; MYRISTYL 19-24; MYRISTYL 61-66; MYRISTYL 121-126; MYRISTYL 125-130; MYRISTYL 46-51; CK2_PHOSPHO_SITE 196-199; MYRISTYL 151-156; MYRISTYL 37-42; PKC_PHOSPHO_SITE 206-208; MYRISTYL 33-38; MYRISTYL 50-55; CK2_PHOSPHO_SITE 206-209;	TGF_BETA_2 145-231; sp_Q64280_TGF4_MOUSE 145-230; TGF-beta 137-231; TGF_BETA_1 159-174; TGFB 141-231;
DEX0449_037.aa.1	N	0 - o1-204;	152-164,1.099; 138-144,1.057; 194-201,1.064; 4-21,1.091; 30-61,1.19; 125-131,1.074; 63-113,1.187;	CK2_PHOSPHO_SITE 113-116; CK2_PHOSPHO_SITE 28-31; MYRISTYL 153-158; MYRISTYL 134-139; CK2_PHOSPHO_SITE 109-112; ASN_GLYCOSYLATION 59-62; MYRISTYL 9-14; PKC_PHOSPHO_SITE 163-165; CK2_PHOSPHO_SITE 141-144;	
DEX0449_038.aa.1	Y	0 - o1-80;	4-36,1.28; 67-77,1.06;	CK2_PHOSPHO_SITE 64-67; MYRISTYL 71-76;	
DEX0449_038.orf.2	N	0 - o1-84;	15-21,1.024; 65-71,1.084; 29-37,1.061;	CK2_PHOSPHO_SITE 59-62; PKC_PHOSPHO_SITE 12-14; PKC_PHOSPHO_SITE 41-43; CK2_PHOSPHO_SITE 60-63:	

				CK2_PHOSPHO_SITE 5-8; CAMP_PHOSPHO_SITE 74-77; PKC_PHOSPHO_SITE 22-24;	
DEX0449_038.aa.3	Y	0 - o1-56;	4-17,1.159; 34-53,1.146;		
DEX0449_039.aa.1	Y	0 - o1-451;	26-42,1.143; 70-77,1.072; 4-11,1.207; 410-415,1.049; 181-193,1.095; 88-100,1.087; 368-381,1.12; 328-341,1.121; 62-68,1.081; 200-260,1.115; 16-24,1.105; 46-60,1.115; 305-311,1.064; 264-277,1.08; 433-439,1.075; 112-118,1.13;	MYRISTYL 351-356; MYRISTYL 365-370; CK2_PHOSPHO_SITE 82-85; CAMP_PHOSPHO_SITE 360-363; PKC_PHOSPHO_SITE 389-391; CK2_PHOSPHO_SITE 336-339; MYRISTYL 415-420; MYRISTYL 106-111; CK2_PHOSPHO_SITE 160-163; PKC_PHOSPHO_SITE 322-324; CK2_PHOSPHO_SITE 125-128; CK2_PHOSPHO_SITE 20-23; MYRISTYL 254-259; CK2_PHOSPHO_SITE 414-417; CK2_PHOSPHO_SITE 294-297; CK2_PHOSPHO_SITE 96-99; LEUCINE_ZIPPER 232-253; ASN_GLYCOSYLATION 158-161; TYR_PHOSPHO_SITE 244-251; MYRISTYL 348-353; CK2_PHOSPHO_SITE 166-169; CK2_PHOSPHO_SITE 129-132; MYRISTYL 438-443; CK2_PHOSPHO_SITE 143-146; AMIDATION 75-78; MYRISTYL 407-412; MYRISTYL 349-354; ASN_GLYCOSYLATION 164-167; CK2_PHOSPHO_SITE 261-264; PKC_PHOSPHO_SITE 359-361;	TPR 45-78; TPR 45-78; TPR 247-280; TPR_REPEAT_1 45-78; TPR_REPEAT_2 205-238; TPR 247-280; TPR 205-238; GLU_RICH 113-321; ATP_GTP_A 89-96; TPR 205-238; TPR_REGION 205-280; TPR_REPEAT_3 247-280;
DEX0449_04	N	0 - o1-	189-	CK2_PHOSPHO_SITE	gmd 24-331;

0.aa.1		337;	208,1.136; 4-11,1.092; 65-75,1.046; 291-301,1.06; 313-319,1.127; 251-260,1.089; 235-249,1.1; 159-165,1.066; 89-100,1.17; 24-30,1.112; 218-224,1.061; 279-289,1.183; 172-178,1.061; 36-57,1.131; 107-157,1.134;	292-295; MYRISTYL 198-203; CK2_PHOSPHO_SITE 32-35; PKC_PHOSPHO_SITE 277-279; MYRISTYL 143-148; PKC_PHOSPHO_SITE 62-64; TYR_PHOSPHO_SITE 115-123; PKC_PHOSPHO_SITE 10-12; CAMP_PHOSPHO_SITE 186-189; CAMP_PHOSPHO_SITE 55-58; MYRISTYL 13-18; PKC_PHOSPHO_SITE 132-134; CK2_PHOSPHO_SITE 154-157; PKC_PHOSPHO_SITE 245-247; MYRISTYL 206-211; MYRISTYL 33-38; CK2_PHOSPHO_SITE 245-248; PKC_PHOSPHO_SITE 185-187; MYRISTYL 36-41; PKC_PHOSPHO_SITE 147-149;	
DEX0449_040.aa.2	N	0 - o1-332;	24-30,1.112; 314-324,1.183; 36-57,1.131; 172-194,1.171; 253-259,1.061; 107-157,1.134; 4-11,1.092; 286-295,1.089; 159-165,1.066; 224-243,1.136; 65-75,1.046; 89-100,1.17; 270-284,1.1; 196-209,1.113;	CK2_PHOSPHO_SITE 32-35; MYRISTYL 204-209; CAMP_PHOSPHO_SITE 55-58; PKC_PHOSPHO_SITE 147-149; MYRISTYL 241-246; TYR_PHOSPHO_SITE 115-123; PKC_PHOSPHO_SITE 62-64; CK2_PHOSPHO_SITE 280-283; PKC_PHOSPHO_SITE 10-12; PKC_PHOSPHO_SITE 132-134; MYRISTYL 143-148; CK2_PHOSPHO_SITE 154-157; MYRISTYL 13-18; PKC_PHOSPHO_SITE 211-213; MYRISTYL 233-238; PKC PHOSPHO SITE	ADH_SHORT 166-194; gmd 24-332;

				312-314; CAMP_PHOSPHO_SITE 221-224; MYRISTYL 33-38; PKC_PHOSPHO_SITE 280-282; CK2_PHOSPHO_SITE 327-330; MYRISTYL 36-41; PKC_PHOSPHO_SITE 220-222;	
DEX0449_040.aa.3	N	0 - o1-382;	172-194,1.171; 196-209,1.113; 286-295,1.089; 159-165,1.066; 253-259,1.061; 314-324,1.183; 343-352,1.075; 224-243,1.136; 107-157,1.134; 36-57,1.131; 65-75,1.046; 4-11,1.092; 24-30,1.112; 357-374,1.131; 326-341,1.108; 89-100,1.17; 270-284,1.1;	PKC_PHOSPHO_SITE 147-149; PKC_PHOSPHO_SITE 132-134; CK2_PHOSPHO_SITE 32-35; MYRISTYL 33-38; PKC_PHOSPHO_SITE 10-12; MYRISTYL 204-209; PKC_PHOSPHO_SITE 280-282; CK2_PHOSPHO_SITE 154-157; PKC_PHOSPHO_SITE 62-64; MYRISTYL 352-357; MYRISTYL 13-18; CK2_PHOSPHO_SITE 280-283; PKC_PHOSPHO_SITE 220-222; TYR_PHOSPHO_SITE 115-123; MYRISTYL 143-148; CAMP_PHOSPHO_SITE 55-58; MYRISTYL 241-246; PKC_PHOSPHO_SITE 211-213; PKC_PHOSPHO_SITE 312-314; CAMP_PHOSPHO_SITE 221-224; MYRISTYL 233-238; MYRISTYL 36-41;	gmd 24-361; ADH_SHORT 166-194;
DEX0449_041.aa.1	N	0 - o1-258;	41-48,1.083; 245-255,1.126; 147-154,1.063; 134-139,1.059; 108-124,1.184; 92-99,1.184; 6-23,1.154; 212-	PKC_PHOSPHO_SITE 232-234; MYRISTYL 159-164; AMIDATION 232-235; CAMP_PHOSPHO_SITE 241-244; CK2_PHOSPHO_SITE 178-181; MYRISTYL 174-179; PKC_PHOSPHO_SITE 115-117:	UQ_con 9-154; sp_Q9BTC1_Q9BTC1_HUMAN 28-153; UBCc 14-157; UBIQUITIN_CONJUGAT_1 84-98; UBIQUITIN_CONJUGAT_2 14-146;

			220,1.139; 61-69,1.073; 168-174,1.029; 189-197,1.112; 71-88,1.083;	CK2_PHOSPHO_SITE 173-176; MYRISTYL 52-57; PKC_PHOSPHO_SITE 162-164; AMIDATION 185-188; CAMP_PHOSPHO_SITE 205-208; MYRISTYL 247-252; PKC_PHOSPHO_SITE 240-242; CK2_PHOSPHO_SITE 3-6; AMIDATION 203-206; PKC_PHOSPHO_SITE 244-246; ASN_GLYCOSYLATION 225-228;	
DEX0449_041.aa.2	N	0 - o1-262;	226-232,1.029; 166-182,1.184; 23-30,1.141; 205-212,1.063; 192-197,1.059; 99-106,1.083; 65-81,1.154; 129-146,1.083; 47-60,1.082; 119-127,1.073; 4-18,1.094; 150-157,1.184;	MYRISTYL 61-66; CK2_PHOSPHO_SITE 231-234; MYRISTYL 248-253; AMIDATION 31-34; MYRISTYL 232-237; PKC_PHOSPHO_SITE 173-175; MYRISTYL 110-115; MYRISTYL 247-252; AMIDATION 41-44; CK2_PHOSPHO_SITE 236-239; MYRISTYL 217-222; MYRISTYL 45-50; MYRISTYL 24-29; MYRISTYL 38-43; PKC_PHOSPHO_SITE 220-222; MYRISTYL 244-249; MYRISTYL 254-259;	sp_Q9BTC1_Q9BTC1_HUMAN 86-211; UBCc 72-215; UBIQUITIN_CONJUGAT_2 72-204; UBIQUITIN_CONJUGAT_1 142-156; UQ_con 67-212;
DEX0449_041.orf.2	N	0 - o1-291;	140-157,1.083; 34-41,1.141; 177-193,1.184; 58-71,1.082; 76-92,1.154; 110-117,1.083; 216-223,1.063; 161-168,1.184; 12-29,1.094; 130-138,1.073; 237-243,1.029; 203-	MYRISTYL 255-260; MYRISTYL 56-61; CK2_PHOSPHO_SITE 247-250; PKC_PHOSPHO_SITE 231-233; PKC_PHOSPHO_SITE 282-284; MYRISTYL 49-54; CK2_PHOSPHO_SITE 242-245; MYRISTYL 72-77; PKC_PHOSPHO_SITE 184-186; AMIDATION 52-55; MYRISTYL 121-126; AMIDATION 42-45; MYRISTYL 35-40; MYRISTYL 243-248;	UBCc 83-226; UQ_con 78-223; sp_Q9BTC1_Q9BTC1_HUMAN 97-222; ARG_RICH 5-59; UBIQUITIN_CONJUGAT_1 153-167; UBIQUITIN_CONJUGAT_2 83-215;

			208,1.059;	MYRISTYL 228-233; CAMP_PHOSPHO_SITE 279-282;	
DEX0449_041.aa.3	N	0 - o1-233;	163-168,1.059; 6-23,1.154; 41-48,1.083; 197-203,1.029; 92-99,1.184; 140-153,1.184; 176-183,1.063; 108-117,1.18; 61-69,1.073; 71-88,1.083; 119-137,1.121;	MYRISTYL 218-223; PKC_PHOSPHO_SITE 116-118; CK2_PHOSPHO_SITE 207-210; MYRISTYL 215-220; CK2_PHOSPHO_SITE 202-205; PKC_PHOSPHO_SITE 144-146; CK2_PHOSPHO_SITE 3-6; MYRISTYL 52-57; MYRISTYL 219-224; CK2_PHOSPHO_SITE 133-136; PKC_PHOSPHO_SITE 191-193; MYRISTYL 188-193; CK2_PHOSPHO_SITE 116-119; MYRISTYL 203-208; MYRISTYL 225-230;	UBIQUITIN_CONJUGAT_2 14-175; UBIQUITIN_CONJUGAT_1 84-98; sp_Q9BTC1_Q9BTC1_HUMAN 28-182; UBCc 14-186; UQ_con 9-183;
DEX0449_041.orf.3	N	0 - o1-390;	145-162,1.154; 279-292,1.184; 336-342,1.029; 180-187,1.083; 315-322,1.063; 231-238,1.184; 200-208,1.073; 247-256,1.18; 15-31,1.195; 41-51,1.11; 65-90,1.129; 210-227,1.083; 258-276,1.121; 302-307,1.059;	MYRISTYL 354-359; AMIDATION 80-83; PKC_PHOSPHO_SITE 283-285; CK2_PHOSPHO_SITE 346-349; CK2_PHOSPHO_SITE 255-258; PKC_PHOSPHO_SITE 255-257; MYRISTYL 128-133; CK2_PHOSPHO_SITE 142-145; MYRISTYL 342-347; AMIDATION 118-121; CAMP_PHOSPHO_SITE 90-93; MYRISTYL 191-196; CAMP_PHOSPHO_SITE 378-381; PKC_PHOSPHO_SITE 381-383; PKC_PHOSPHO_SITE 33-35; MYRISTYL 85-90; CK2_PHOSPHO_SITE 272-275; PKC_PHOSPHO_SITE 45-47; PKC_PHOSPHO_SITE 330-332; CK2_PHOSPHO_SITE	OXYTOCINR 324-337; UBCc 153-325; UBIQUITIN_CONJUGAT_2 153-314; UBIQUITIN_CONJUGAT_1 223-237; UQ_con 148-322; sp_Q9BTC1_Q9BTC1_HUMAN 167-321; OXYTOCINR 93-106;

				341-344; MYRISTYL 327-332;	
DEX0449_042.aa.1	N	0 - o1-110;	59-66,1.064; 23-35,1.2; 87-107,1.135; 43-51,1.055; 4-19,1.144; 71-80,1.11;	PKC_PHOSPHO_SITE 92-94; CAMP_PHOSPHO_SITE 37-40; MYRISTYL 2-7; PKC_PHOSPHO_SITE 35-37; MYRISTYL 56-61;	
DEX0449_043.aa.1	N	1 - i1-95;tm96-118;o119-257;	48-55,1.06; 228-234,1.108; 249-254,1.068; 202-210,1.12; 129-146,1.215; 62-72,1.097; 212-224,1.14; 157-169,1.121; 185-191,1.092; 94-118,1.175; 76-82,1.04; 171-180,1.089; 240-247,1.039; 22-34,1.191;	MYRISTYL 9-14; PKC_PHOSPHO_SITE 214-216; PKC_PHOSPHO_SITE 148-150; CK2_PHOSPHO_SITE 210-213; PKC_PHOSPHO_SITE 184-186; CK2_PHOSPHO_SITE 201-204; PKC_PHOSPHO_SITE 224-226;	AhpC-TSA 67-216;
DEX0449_044.aa.1	N	0 - i1-105;	74-81,1.122; 95-102,1.095; 24-48,1.119;	CAMP_PHOSPHO_SITE 4-7; AMIDATION 16-19; CAMP_PHOSPHO_SITE 63-66; PKC_PHOSPHO_SITE 91-93; ASN_GLYCOSYLATION 21-24; ASN_GLYCOSYLATION 56-59; PKC_PHOSPHO_SITE 69-71; ASN_GLYCOSYLATION 65-68; PKC_PHOSPHO_SITE 58-60;	sp_P02403_RL37_HUMAN 12-62; Ribosomal_L37e 10-63; RIBOSOMAL_L37E 12-31;
DEX0449_044.orf.1	N	0 - o1-111;	30-54,1.119; 80-87,1.122; 102-108,1.095;	PKC_PHOSPHO_SITE 75-77; ASN_GLYCOSYLATION 71-74; ASN_GLYCOSYLATION 27-30; AMIDATION 22-25:	Ribosomal_L37e 18-69; sp_P02403_RL37_HUMAN 18-68; RIBOSOMAL_L37E 18-37;

				ASN_GLYCOSYLATION 62-65; PKC_PHOSPHO_SITE 64-66; CAMP_PHOSPHO_SITE 69-72; PKC_PHOSPHO_SITE 97-99;	
DEX0449_044.aa.2	N	0 - o1-104;		ASN_GLYCOSYLATION 20-23; PKC_PHOSPHO_SITE 2-4; ASN_GLYCOSYLATION 55-58; PKC_PHOSPHO_SITE 57-59; AMIDATION 15-18; CK2_PHOSPHO_SITE 2-5; CAMP_PHOSPHO_SITE 62-65; PKC_PHOSPHO_SITE 4-6; PKC_PHOSPHO_SITE 90-92; PKC_PHOSPHO_SITE 68-70; ASN_GLYCOSYLATION 54-57;	Ribosomal_L37e 9-62; sp_P02403_RL37_HUMAN 10-61; RIBOSOMAL_L37E 11-30;
DEX0449_044.orf.2	N	0 - o1-129;		MYRISTYL 31-36; PKC_PHOSPHO_SITE 94-96; PKC_PHOSPHO_SITE 2-4; MYRISTYL 50-55; ASN_GLYCOSYLATION 35-38; PKC_PHOSPHO_SITE 108-110; ASN_GLYCOSYLATION 12-15; PKC_PHOSPHO_SITE 11-13; CK2_PHOSPHO_SITE 62-65; CK2_PHOSPHO_SITE 108-111; PKC_PHOSPHO_SITE 20-22; PKC_PHOSPHO_SITE 10-12;	
DEX0449_044.aa.3	N	0 - i1-96;	15-39,1.104; 87-93,1.095; 4-9,1.129; 65-72,1.122;	ASN_GLYCOSYLATION 56-59; CAMP_PHOSPHO_SITE 54-57; ASN_GLYCOSYLATION 47-50; PKC_PHOSPHO_SITE 82-84;	Ribosomal_L37e 7-54; sp_P02403_RL37_HUMAN 12-53;

				PKC_PHOSPHO_SITE 60-62; PKC_PHOSPHO_SITE 49-51;	
DEX0449_044.aa.4	N	0 - i1-55;	23-36,1.118; 39-45,1.072;	PKC_PHOSPHO_SITE 20-22; CK2_PHOSPHO_SITE 34-37; PKC_PHOSPHO_SITE 34-36; CK2_PHOSPHO_SITE 7-10;	
DEX0449_044.orf.4	N	0 - o1-72;	63-69,1.095; 4-15,1.104; 41-48,1.122;	PKC_PHOSPHO_SITE 58-60; CAMP_PHOSPHO_SITE 30-33; PKC_PHOSPHO_SITE 25-27; PKC_PHOSPHO_SITE 36-38; ASN_GLYCOSYLATION 23-26; ASN_GLYCOSYLATION 32-35;	Ribosomal_L37e 1-30; sp_Q9D823_Q9D823_MOUSE 1-29;
DEX0449_045.aa.1	N	0 - o1-185;	7-13,1.077; 21-27,1.045; 65-72,1.156; 42-56,1.188; 137-162,1.156; 169-175,1.06; 100-106,1.164; 110-127,1.193;	ASN_GLYCOSYLATION 98-101; MYRISTYL 9-14; MYRISTYL 178-183; ASN_GLYCOSYLATION 61-64; CK2_PHOSPHO_SITE 106-109; PKC_PHOSPHO_SITE 151-153; MYRISTYL 14-19; PKC_PHOSPHO_SITE 127-129; PKC_PHOSPHO_SITE 77-79; MYRISTYL 19-24;	
DEX0449_045.orf.1	N	0 - o1-214;	194-211,1.128; 14-21,1.059; 182-191,1.098; 86-101,1.148; 49-67,1.235; 72-83,1.199; 105-113,1.105; 120-180,1.17; 29-47,1.107; 4-11,1.148;	MYRISTYL 71-76; ASN_GLYCOSYLATION 147-150; MYRISTYL 145-150; PKC_PHOSPHO_SITE 176-178; MYRISTYL 150-155; PKC_PHOSPHO_SITE 66-68; PKC_PHOSPHO_SITE 34-36;	
DEX0449_045.aa.2	N	0 - o1-131;	7-13,1.077; 21-27,1.045; 100-	ASN_GLYCOSYLATION 61-64; ASN GLYCOSYLATION	RIBOSOMAL_L18E 70-87; sp 007020 RL18 HUMA

			106,1.164; 65-72,1.166; 42-56,1.188; 110-122,1.136;	96-101; MYRISTYL 19-24; CK2_PHOSPHO_SITE 106-109; PKC_PHOSPHO_SITE 77-79; MYRISTYL 9-14; MYRISTYL 14-19;	N 49-119; Ribosomal_L18e 31-131;
DEX0449_045.orf.2	N	0 - ol-132;	22-28,1.045; 101-107,1.164; 66-73,1.166; 9-14,1.005; 43-57,1.188; 111-123,1.136;	ASN_GLYCOSYLATION 62-65; ASN_GLYCOSYLATION 99-102; MYRISTYL 15-20; CK2_PHOSPHO_SITE 107-110; MYRISTYL 20-25; MYRISTYL 10-15; MYRISTYL 6-11; PKC_PHOSPHO_SITE 78-80;	RIBOSOMAL_L18E 71-88; Ribosomal_L18e 32-132; sp_Q07020_RL18_HUMAN 50-120;
DEX0449_045.aa.3	N	0 - ol-195;	180-191,1.088; 143-153,1.193; 74-80,1.087; 14-20,1.045; 58-65,1.166; 94-100,1.115; 35-49,1.188; 160-172,1.133; 115-141,1.177;	ASN_GLYCOSYLATION 54-57; MYRISTYL 7-12; MYRISTYL 172-177; MYRISTYL 127-132; MYRISTYL 12-17; MYRISTYL 109-114; MYRISTYL 189-194; PKC_PHOSPHO_SITE 4-6; PKC_PHOSPHO_SITE 70-72;	sp_Q9D987_Q9D987_MOUSE 42-81; Ribosomal_L18e 24-160; RIBOSOMAL_L18E 63-80;
DEX0449_046.aa.1	Y	0 - ol-115;	18-51,1.229; 54-109,1.225;	PKC_PHOSPHO_SITE 107-109; MYRISTYL 34-39; MYRISTYL 102-107; PKC_PHOSPHO_SITE 38-40; MYRISTYL 98-103; TYR_PHOSPHO_SITE 40-48;	
DEX0449_046.aa.2	N	0 - ol-156;	91-126,1.217; 37-83,1.178; 133-149,1.145;	MYRISTYL 126-131; MYRISTYL 6-11; CK2_PHOSPHO_SITE 30-33; ASN_GLYCOSYLATION 28-31; MYRISTYL 95-100; MYRISTYL 38-43; PKC_PHOSPHO_SITE 150-152; PKC_PHOSPHO_SITE 9-11; MYRISTYL 17-22; MYRISTYL 14-19; PKC PHOSPHO SITE	TIMP 13-148; NTR 26-148;

				127-129;	.
DEX0449_047.aa.1	Y	0 - o1-226;	141-154,1.109; 159-176,1.17; 180-210,1.214; 42-57,1.155; 212-223,1.121; 76-133,1.186; 4-22,1.064; 32-40,1.093;	MYRISTYL 26-31; CK2_PHOSPHO_SITE 71-74; MYRISTYL 21-26; MYRISTYL 9-14; CK2_PHOSPHO_SITE 132-135; CK2_PHOSPHO_SITE 73-76; MYRISTYL 27-32; CK2_PHOSPHO_SITE 67-70; MYRISTYL 34-39; MYRISTYL 44-49;	CYCLIN 83-185; KV14CHANNEL 19-28; ALA_RICH 2-22; ANTIFREEZEI 13-22; ANTIFREEZEI 1-11; KV14CHANNEL 6-16; cyclin 48-192;
DEX0449_047.orf.1	N	0 - o1-253;	103-125,1.18; 141-154,1.056; 4-13,1.152; 176-182,1.102; 158-167,1.11; 37-48,1.194; 57-88,1.216; 90-96,1.106; 15-24,1.127; 197-203,1.06;	PKC_PHOSPHO_SITE 229-231; CK2_PHOSPHO_SITE 167-170; CAMP_PHOSPHO_SITE 232-235; PKC_PHOSPHO_SITE 184-186; ASN_GLYCOSYLATION 53-56; MYRISTYL 143-148; MYRISTYL 30-35; PKC_PHOSPHO_SITE 180-182; TYR_PHOSPHO_SITE 214-220; PKC_PHOSPHO_SITE 11-13; MYRISTYL 202-207; PKC_PHOSPHO_SITE 246-248; CK2_PHOSPHO_SITE 98-101; MYRISTYL 238-243; PKC_PHOSPHO_SITE 167-169; MYRISTYL 242-247; CK2_PHOSPHO_SITE 121-124; PKC_PHOSPHO_SITE 55-57; MYRISTYL 190-195;	SER_RICH 209-249; CYCLIN 13-111;
DEX0449_047.orf.2	N	0 - o1-346;	57-88,1.216; 176-182,1.102; 103-125,1.18; 15-24,1.127; 197-203,1.06; 4-13,1.152; 37-48,1.194; 280-287.1.082:	PKC_PHOSPHO_SITE 180-182; TYR_PHOSPHO_SITE 276-283; PKC_PHOSPHO_SITE 333-335; MYRISTYL 190-195; PKC_PHOSPHO_SITE 275-277; PKC_PHOSPHO_SITE 295-297; MYRISTYL 242-247: MYRISTYL	HIGHMOBLTYIY 249-268; SER_RICH 209-302; HIGHMOBLTYIY 222-233; TUBBYPROTEIN 171-196; TUBBYPROTEIN 285-311; HIGHMOBLTYIY 207-219; ARG_RICH 207-339; CYCLIN 13-111;

			315-320,1.033; 158-167,1.11; 141-154,1.056; 90-96,1.106;	30-35; PKC_PHOSPHO_SITE 184-186; CK2_PHOSPHO_SITE 98-101; MYRISTYL 238-243; CK2_PHOSPHO_SITE 121-124; CAMP_PHOSPHO_SITE 232-235; CAMP_PHOSPHO_SITE 313-316; CK2_PHOSPHO_SITE 167-170; MYRISTYL 274-279; AMIDATION 333-336; PKC_PHOSPHO_SITE 229-231; PKC_PHOSPHO_SITE 11-13; MYRISTYL 143-148; ASN_GLYCOSYLATION 53-56; PKC_PHOSPHO_SITE 167-169; MYRISTYL 202-207; TYR_PHOSPHO_SITE 214-220; PKC_PHOSPHO_SITE 55-57; CK2_PHOSPHO_SITE 252-255;	
DEX0449_048.aa.1	Y	2 - i1-4;tm5-27;o28-52;tm53-75;i76-91;	63-85,1.21; 40-61,1.119; 13-28,1.179; 6-11,1.065;	MYRISTYL 59-64; PKC_PHOSPHO_SITE 3-5; PKC_PHOSPHO_SITE 47-49; PKC_PHOSPHO_SITE 26-28;	CD225 1-67;
DEX0449_048.orf.1	N	0 - i1-92;	12-52,1.184; 59-89,1.161;	MYRISTYL 40-45; MYRISTYL 1-6; MYRISTYL 59-64; PKC_PHOSPHO_SITE 55-57; TYR_PHOSPHO_SITE 42-49;	
DEX0449_049.aa.1	N	0 - o1-137;	66-73,1.106; 27-39,1.189; 89-95,1.055; 118-126,1.081; 128-134,1.045; 41-61,1.185; 75-83,1.108;	CK2_PHOSPHO_SITE 107-110; PKC_PHOSPHO_SITE 29-31; PKC_PHOSPHO_SITE 64-66; MYRISTYL 39-44; CK2_PHOSPHO_SITE 6-9; CK2_PHOSPHO_SITE 62-65; PKC_PHOSPHO_SITE	Ribophorin_I 2-137;

				24-26; PKC_PHOSPHO_SITE 128-130; TYR_PHOSPHO_SITE 108-115; PKC_PHOSPHO_SITE 87-89;	
DEX0449_049.orf.1	N	0 - ol-148;		MYRISTYL 50-55; PKC_PHOSPHO_SITE 98-100; PKC_PHOSPHO_SITE 35-37; CK2_PHOSPHO_SITE 118-121; PKC_PHOSPHO_SITE 139-141; CK2_PHOSPHO_SITE 22-25; TYR_PHOSPHO_SITE 119-126; CK2_PHOSPHO_SITE 17-20; TYR_PHOSPHO_SITE 16-24; CK2_PHOSPHO_SITE 73-76; PKC_PHOSPHO_SITE 40-42; PKC_PHOSPHO_SITE 75-77;	Ribophorin_I 3-148;
DEX0449_050.aa.1	Y	0 - ol-479;	466-476,1.085; 230-237,1.111; 394-403,1.1; 126-146,1.185; 5-17,1.227; 151-158,1.106; 253-261,1.094; 405-448,1.182; 160-168,1.108; 300-311,1.133; 174-180,1.055; 67-76,1.118; 328-376,1.147; 37-59,1.187; 203-211,1.081; 112-124,1.189; 280-	PKC_PHOSPHO_SITE 369-371; CK2_PHOSPHO_SITE 470-473; MYRISTYL 124-129; MYRISTYL 298-303; CK2_PHOSPHO_SITE 453-456; MYRISTYL 59-64; TYR_PHOSPHO_SITE 391-398; CK2_PHOSPHO_SITE 399-402; PKC_PHOSPHO_SITE 63-65; AMIDATION 462-465; PKC_PHOSPHO_SITE 276-278; MYRISTYL 22-27; PKC_PHOSPHO_SITE 213-215; PKC_PHOSPHO_SITE 114-116; CK2_PHOSPHO_SITE 437-440; MYRISTYL 60-65; PKC_PHOSPHO_SITE 149-151; CK2 PHOSPHO SITE	Ribophorin_I 30-437;

			294,1.102; 26-34,1.065; 78-89,1.153;	192-195; ASN_GLYCOSYLATION 299-302; CAMP_PHOSPHO_SITE 450-453; CK2_PHOSPHO_SITE 147-150; PKC_PHOSPHO_SITE 109-111; CK2_PHOSPHO_SITE 385-388; CK2_PHOSPHO_SITE 262-265; PKC_PHOSPHO_SITE 462-464; PKC_PHOSPHO_SITE 279-281; MYRISTYL 324-329; TYR_PHOSPHO_SITE 193-200; PKC_PHOSPHO_SITE 172-174;	
DEX0449_050.aa.2	Y	0 - o1-361;	160-168,1.108; 67-76,1.118; 37-59,1.187; 230-237,1.111; 280-294,1.102; 112-124,1.189; 328-354,1.108; 26-34,1.065; 5-17,1.227; 300-311,1.133; 203-211,1.081; 151-158,1.106; 126-146,1.185; 253-261,1.094; 174-180,1.055; 78-89,1.153;	MYRISTYL 124-129; TYR_PHOSPHO_SITE 193-200; PKC_PHOSPHO_SITE 149-151; MYRISTYL 324-329; MYRISTYL 22-27; PKC_PHOSPHO_SITE 172-174; CK2_PHOSPHO_SITE 147-150; MYRISTYL 298-303; PKC_PHOSPHO_SITE 114-116; PKC_PHOSPHO_SITE 276-278; CK2_PHOSPHO_SITE 192-195; PKC_PHOSPHO_SITE 63-65; PKC_PHOSPHO_SITE 213-215; ASN_GLYCOSYLATION 299-302; PKC_PHOSPHO_SITE 109-111; MYRISTYL 60-65; PKC_PHOSPHO_SITE 279-281; MYRISTYL 59-64; CK2_PHOSPHO_SITE 262-265;	Ribophorin_I 30-361;
DEX0449_050.aa.3	Y	0 - o1-420;	395-406,1.196; 26-34,1.065; 37-59,1.187; 5-17,1.227;	MYRISTYL 124-129; MYRISTYL 22-27; PKC_PHOSPHO_SITE 410-412; MYRISTYL 377-382;	Ribophorin_I 30-412;

			67-76,1.118; 174-180,1.055; 126-146,1.185; 410-417,1.041; 230-237,1.111; 151-158,1.106; 300-311,1.133; 203-211,1.081; 160-168,1.108; 78-89,1.153; 328-376,1.147; 253-261,1.094; 280-294,1.102; 112-124,1.189; 383-393,1.162;	PKC_PHOSPHO_SITE 172-174; CK2_PHOSPHO_SITE 192-195; PKC_PHOSPHO_SITE 109-111; MYRISTYL 394-399; MYRISTYL 59-64; PKC_PHOSPHO_SITE 114-116; CK2_PHOSPHO_SITE 262-265; TYR_PHOSPHO_SITE 193-200; PKC_PHOSPHO_SITE 149-151; ASN_GLYCOSYLATION 299-302; MYRISTYL 60-65; PKC_PHOSPHO_SITE 276-278; CK2_PHOSPHO_SITE 147-150; MYRISTYL 324-329; PKC_PHOSPHO_SITE 279-281; PKC_PHOSPHO_SITE 63-65; MYRISTYL 298-303; PKC_PHOSPHO_SITE 213-215; PKC_PHOSPHO_SITE 369-371;	
DEX0449_051.aa.1	N	0 - o1-128;	8-13,1.039; 68-76,1.059; 49-64,1.126; 104-117,1.058;	MYRISTYL 33-38; MYRISTYL 116-121; PKC_PHOSPHO_SITE 106-108; PKC_PHOSPHO_SITE 24-26; MYRISTYL 44-49; MYRISTYL 91-96; CK2_PHOSPHO_SITE 95-98; MYRISTYL 14-19; MYRISTYL 112-117;	
DEX0449_052.aa.1	N	0 - o1-315;	170-187,1.133; 93-115,1.194; 256-263,1.083; 35-54,1.172; 58-81,1.224; 294-309,1.089; 267-273,1.122; 17-30,1.123; 199-	CK2_PHOSPHO_SITE 160-163; MYRISTYL 81-86; ASN_GLYCOSYLATION 155-158; MYRISTYL 136-141; CK2_PHOSPHO_SITE 245-248; MYRISTYL 115-120; PKC_PHOSPHO_SITE 21-23; CK2_PHOSPHO_SITE 15-18; MYRISTYL 5-10;	ank 124-156; ANK 124-153; ank 57-89; ank 91-123; ANKYRIN 137-149; ANKYRIN 125-137; ANK_REP_REGION 21-179; ANK 21-50; ank 21-55; ANK_REPEAT_3 124-156; ANK_REPEAT_2 91-123; ANK_REPEAT_1 57-89; ank 157-190; ANK 57-86; ANK 91-120;

			208,1.063; 83-89,1.046; 159-165,1.053; 137-154,1.191; 233-241,1.127;	ASN_GLYCOSYLATION 85-88; PKC_PHOSPHO_SITE 88-90; MYRISTYL 196-201; CK2_PHOSPHO_SITE 241-244; MYRISTYL 192-197; MYRISTYL 158-163; MYRISTYL 223-228; MYRISTYL 9-14; PKC_PHOSPHO_SITE 210-212;	
DEX0449_053.aa.1	N	0 - o1-142;	40-48,1.074; 64-94,1.145; 108-117,1.063;	CAMP_PHOSPHO_SITE 17-20; PKC_PHOSPHO_SITE 20-22; PKC_PHOSPHO_SITE 132-134; MYRISTYL 38-43; PKC_PHOSPHO_SITE 137-139; PKC_PHOSPHO_SITE 10-12; CAMP_PHOSPHO_SITE 135-138; PKC_PHOSPHO_SITE 21-23; MYRISTYL 116-121; MYRISTYL 2-7; CK2_PHOSPHO_SITE 93-96; CAMP_PHOSPHO_SITE 134-137; AMIDATION 15-18; AMIDATION 120-123;	
DEX0449_053.aa.2	N	0 - o1-72;	14-32,1.094; 40-46,1.064;	CAMP_PHOSPHO_SITE 65-68; CK2_PHOSPHO_SITE 38-41; CAMP_PHOSPHO_SITE 64-67; MYRISTYL 46-51; PKC_PHOSPHO_SITE 62-64; PKC_PHOSPHO_SITE 67-69; AMIDATION 50-53; MYRISTYL 2-7;	
DEX0449_053.orf.2	N	0 - o1-122;	67-108,1.199; 59-65,1.06; 11-47,1.156;	CK2_PHOSPHO_SITE 1-4; AMIDATION 115-118; ASN_GLYCOSYLATION 100-103; MYRISTYL 13-18; PKC_PHOSPHO_SITE 59-61; MYRISTYL 27-32;	

				ASN_GLYCOSYLATION 107-110; PKC_PHOSPHO_SITE 95-97; MYRISTYL 53-58;	
DEX0449_054.aa.1	N	0 - o1-42;	5-11,1.101; 29-39,1.142;	MYRISTYL 15-20; CK2_PHOSPHO_SITE 23-26; MYRISTYL 11-16; MYRISTYL 20-25; MYRISTYL 6-11; MYRISTYL 16-21;	
DEX0449_054.orf.1	N	0 - o1-260;	54-80,1.148; 187-194,1.044; 197-207,1.183; 86-96,1.133; 34-52,1.074; 118-126,1.089; 14-24,1.051;	CK2_PHOSPHO_SITE 96-99; CK2_PHOSPHO_SITE 145-148; PKC_PHOSPHO_SITE 145-147; CK2_PHOSPHO_SITE 91-94; PKC_PHOSPHO_SITE 135-137; ASN_GLYCOSYLATION 129-132; PKC_PHOSPHO_SITE 161-163; CAMP_PHOSPHO_SITE 151-154; PKC_PHOSPHO_SITE 150-152;	GLU_RICH 211-258; ASP_RICH 176-260; NAP 12-209;
DEX0449_054.orf.2	N	0 - o1-285;	111-121,1.133; 212-219,1.044; 79-105,1.148; 4-33,1.166; 39-45,1.051; 222-232,1.183; 59-77,1.074; 143-151,1.089;	PKC_PHOSPHO_SITE 170-172; PKC_PHOSPHO_SITE 160-162; PKC_PHOSPHO_SITE 186-188; ASN_GLYCOSYLATION 154-157; CK2_PHOSPHO_SITE 121-124; CK2_PHOSPHO_SITE 1-4; CK2_PHOSPHO_SITE 170-173; CAMP_PHOSPHO_SITE 176-179; PKC_PHOSPHO_SITE 175-177; CK2_PHOSPHO_SITE 116-119;	ASP_RICH 203-285; NAP 37-234; GLU_RICH 236-283;
DEX0449_055.aa.1	N	0 - o1-600;	462-469,1.068; 36-47,1.086; 96-102,1.051; 494-506,1.08; 263-	CK2_PHOSPHO_SITE 309-312; CK2_PHOSPHO_SITE 245-248; ASN_GLYCOSYLATION 112-115; MYRISTYL 515-520; CK2_PHOSPHO_SITE	IF 368-376; IF_tail 422-559; PRENYLATION 597-600; filament 25-381;

			272,1.055; 170-176,1.091; 248-261,1.123; 541-547,1.129; 189-198,1.061; 427-434,1.067; 219-227,1.108; 508-530,1.101; 13-19,1.061; 295-309,1.102; 63-89,1.053; 52-58,1.102; 154-166,1.071; 69-79,1.095; 348-365,1.142; 484-490,1.068; 441-458,1.086; 382-397,1.091; 122-148,1.132;	59-62; PKC_PHOSPHO_SITE 408-410; ASN_GLYCOSYLATION 470-473; MYRISTYL 433-438; TYR_PHOSPHO_SITE 291-299; MYRISTYL 435-440; PKC_PHOSPHO_SITE 226-228; PKC_PHOSPHO_SITE 65-67; PKC_PHOSPHO_SITE 472-474; PKC_PHOSPHO_SITE 498-500; PKC_PHOSPHO_SITE 148-150; PKC_PHOSPHO_SITE 114-116; TYR_PHOSPHO_SITE 255-262; CK2_PHOSPHO_SITE 204-207; PKC_PHOSPHO_SITE 90-92; PKC_PHOSPHO_SITE 414-416; MYRISTYL 403-408; MYRISTYL 426-431; PKC_PHOSPHO_SITE 389-391; PKC_PHOSPHO_SITE 556-558; CAMP_PHOSPHO_SITE 558-561; MYRISTYL 506-511; PKC_PHOSPHO_SITE 59-61; CK2_PHOSPHO_SITE 65-68; PKC_PHOSPHO_SITE 268-270; CK2_PHOSPHO_SITE 455-458; CK2_PHOSPHO_SITE 194-197; PKC_PHOSPHO_SITE 213-215; CK2_PHOSPHO_SITE 139-142; MYRISTYL 535-540; CK2_PHOSPHO_SITE 472-475; CK2_PHOSPHO_SITE 90-93; PKC_PHOSPHO_SITE 539-541;	

				PKC_PHOSPHO_SITE 593-595; MYRISTYL 437-442;	
DEX0449_055.orf.1	N	0 - o1-620;	447-454,1.067; 368-385,1.142; 315-329,1.102; 142-168,1.132; 33-39,1.061; 116-122,1.051; 103-109,1.053; 402-417,1.091; 504-510,1.068; 461-478,1.086; 209-218,1.061; 72-78,1.102; 482-489,1.068; 89-99,1.095; 528-550,1.101; 174-186,1.071; 56-67,1.086; 190-196,1.091; 239-247,1.108; 268-281,1.123; 514-526,1.08; 283-292,1.055; 561-567,1.129; 22-27,1.014;	PKC_PHOSPHO_SITE 134-136; PKC_PHOSPHO_SITE 409-411; MYRISTYL 453-458; MYRISTYL 446-451; PKC_PHOSPHO_SITE 576-578; PKC_PHOSPHO_SITE 246-248; CK2_PHOSPHO_SITE 85-88; PKC_PHOSPHO_SITE 288-290; CK2_PHOSPHO_SITE 214-217; MYRISTYL 555-560; CK2_PHOSPHO_SITE 492-495; PKC_PHOSPHO_SITE 168-170; PKC_PHOSPHO_SITE 613-615; PKC_PHOSPHO_SITE 428-430; TYR_PHOSPHO_SITE 275-282; PKC_PHOSPHO_SITE 434-436; CK2_PHOSPHO_SITE 159-162; PKC_PHOSPHO_SITE 518-520; TYR_PHOSPHO_SITE 311-319; CK2_PHOSPHO_SITE 475-478; CAMP_PHOSPHO_SITE 578-581; PKC_PHOSPHO_SITE 492-494; CK2_PHOSPHO_SITE 224-227; ASN_GLYCOSYLATION 490-493; PKC_PHOSPHO_SITE 233-235; ASN_GLYCOSYLATION 132-135; CK2_PHOSPHO_SITE 265-268; MYRISTYL 423-428; CK2_PHOSPHO_SITE 329-332; PKC_PHOSPHO_SITE 79-81;	PRENYLATION 617-620; IF_tail 442-579; IF 388-396; filament 45-401;

				CK2_PHOSPHO_SITE 79-82; MYRISTYL 455-460; PKC_PHOSPHO_SITE 110-112; MYRISTYL 535-540; AMIDATION 6-9; PKC_PHOSPHO_SITE 559-561; PKC_PHOSPHO_SITE 85-87; CK2_PHOSPHO_SITE 110-113; MYRISTYL 526-531; MYRISTYL 457-462;	
DEX0449_055.aa.2	N	0 - o1-237;	174-186,1.071; 33-39,1.061; 199-207,1.122; 218-234,1.134; 103-109,1.053; 56-67,1.086; 142-168,1.132; 89-99,1.095; 22-27,1.014; 72-78,1.102; 190-196,1.091; 116-122,1.051;	CK2_PHOSPHO_SITE 79-82; PKC_PHOSPHO_SITE 110-112; CK2_PHOSPHO_SITE 85-88; ASN_GLYCOSYLATION 132-135; AMIDATION 6-9; PKC_PHOSPHO_SITE 79-81; PKC_PHOSPHO_SITE 85-87; MYRISTYL 201-206; PKC_PHOSPHO_SITE 205-207; PKC_PHOSPHO_SITE 168-170; CK2_PHOSPHO_SITE 110-113; CK2_PHOSPHO_SITE 159-162; PKC_PHOSPHO_SITE 134-136; CK2_PHOSPHO_SITE 210-213;	filament 45-232;
DEX0449_055.orf.3	N	0 - o1-620;	482-489,1.068; 268-281,1.123; 22-27,1.014; 561-567,1.129; 56-67,1.086; 174-186,1.071; 368-385,1.142; 528-550,1.101; 447-454,1.067; 72-78,1.102; 283-	CAMP_PHOSPHO_SITE 578-581; TYR_PHOSPHO_SITE 275-282; CK2_PHOSPHO_SITE 329-332; AMIDATION 6-9; CK2_PHOSPHO_SITE 475-478; PKC_PHOSPHO_SITE 613-615; PKC_PHOSPHO_SITE 85-87; MYRISTYL 446-451; CK2_PHOSPHO_SITE 110-113; MYRISTYL 555-560; CK2 PHOSPHO SITE	PRENYLATION 617-620; IF 386-396; filament 45-401; IF_tail 442-579;

			292,1.055; 209- 218,1.061; 89-99,1.095; 514- 526,1.08; 33-39,1.061; 461- 478,1.086; 504- 510,1.068; 315- 329,1.102; 103- 109,1.053; 239- 247,1.108; 402- 417,1.091; 190- 196,1.091; 142- 168,1.132; 116- 122,1.051;	79-82; CK2_PHOSPHO_SITE 85-88; PKC_PHOSPHO_SITE 559-561; TYR_PHOSPHO_SITE 311-319; PKC_PHOSPHO_SITE 79-81; PKC_PHOSPHO_SITE 168-170; MYRISTYL 457-462; CK2_PHOSPHO_SITE 159-162; PKC_PHOSPHO_SITE 134-136; MYRISTYL 455-460; MYRISTYL 453-458; ASN_GLYCOSYLATION 132-135; CK2_PHOSPHO_SITE 492-495; ASN_GLYCOSYLATION 490-493; PKC_PHOSPHO_SITE 428-430; CK2_PHOSPHO_SITE 214-217; MYRISTYL 526-531; PKC_PHOSPHO_SITE 576-578; PKC_PHOSPHO_SITE 409-411; PKC_PHOSPHO_SITE 288-290; PKC_PHOSPHO_SITE 110-112; PKC_PHOSPHO_SITE 434-436; MYRISTYL 535-540; PKC_PHOSPHO_SITE 233-235; MYRISTYL 423-428; PKC_PHOSPHO_SITE 492-494; CK2_PHOSPHO_SITE 265-268; PKC_PHOSPHO_SITE 518-520; CK2_PHOSPHO_SITE 224-227; PKC_PHOSPHO_SITE 246-248;	

Example 1b: Sequence Alignment Support

Alignments between previously identified sequences and splice variant sequences are performed to confirm unique portions of splice variant nucleic acid and amino acid sequences. The alignments are done using the Needle program in the European Molecular Biology Open Software Suite (EMBOSS) version 2.2.0 available at www.emboss.org from EMBnet (http://www.embnet.org). Default settings are used unless otherwise noted. The Needle program in EMBOSS implements the Needleman-Wunsch algorithm. Needleman, S. B., Wunsch, C. D., *J. Mol. Biol.* 48:443-453 (1970).

It is well know to those skilled in the art that implication of alignment algorithms by various programs may result in minor changes in the generated output. These changes include but are not limited to: alignment scores (percent identity, similarity, and gap), display of nonaligned flanking sequence regions, and number assignment to residues. These minor changes in the output of an alignment do not alter the physical characteristics of the sequences or the differences between the sequences, e.g. regions of homology, insertions, or deletions.

Example 1c: RT-PCR Analysis

To detect the presence and tissue distribution of a particular splice variant Reverse Transcription-Polymerase Chain Reaction (RT-PCR) is performed using cDNA generated from a panel of tissue RNAs. *See, e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and; Kawasaki ES *et al.*, *PNAS* 85(15):5698 (1988). Total RNA is extracted from a variety of tissues and first strand cDNA is prepared with reverse transcriptase (RT). Each panel includes 23 cDNAs from five cancer types (lung, ovary, breast, colon, and prostate) and normal samples of testis, placenta and fetal brain. Each cancer set is composed of three cancer cDNAs from different donors and one normal pooled sample. Using a standard enzyme kit from BD Bioscience Clontech (Mountain View, CA), the target transcript is detected with sequence-specific primers designed to only amplify the particular splice variant. The PCR reaction is run on the GeneAmp PCR system 9700 (Applied Biosystem, Foster City, CA) thermocycler under optimal conditions. One of ordinary skill can design appropriate primers and determine optimal conditions. The amplified product is resolved on an agarose gel to detect a band of equivalent size to the predicted RT-PCR product. A band indicated the presence of the splice variant in a sample. The relation of the amplified product to the splice variant was subsequently confirmed by DNA sequencing.

After subcloning, all positively screened clones are sequence verified. The DNA sequence verification results show the splice variant contains the predicted sequence differences in comparison with the reference sequence.

Results for RT-PCR analysis include the sequence DEX ID, Lead Name, Cancer Tissue(s) the transcript was detected in, Normal Tissue(s) the transcript was detected in, the predicted length of the RT-PCR product, and the Confirmed Length of the RT-PCR product.

RT-PCR results confirm the presence SEQ ID NO: 1-112 in biologic samples and distinguish between related transcripts.

Example 1d: Secretion Assay

To determine if a protein encoded by a splice variant is secreted from cells a secretion assay is preformed. A pcDNA3.1 clone containing the gene transcript which encodes the variant protein is transfected into 293T cells using the Superfect transfection reagent (Qiagen, Valencia CA). Transfected cells are incubated for 28 hours before the media is collected and immediately spun down to remove any detached cells. The adherent cells are solubilized with lysis buffer (1% NP40, 10mM sodium phosphate pH7.0, and 0.15M NaCl). The lysed cells are collected and spun down and the supernatant extracted as cell lysate. Western immunoblot is carried out in the following manner: 15μl of the cell lysate and media are run on 4-12% NuPage Bis-Tris gel (Invitrogen, Carlsbad CA), and blotted onto a PVDF membrane (Invitrogen, Carlsbad CA). The blot is incubated with a polyclonal primary antibody which binds to the variant protein (Imgenex, San Diego CA) and polyclonal goat anti-rabbit-peroxidase secondary antibody (Sigma-Aldrich, St. Louis MO). The blot is developed with the ECL Plus chemiluminescent detection reagent (Amersham BioSciences, Piscataway NJ).

Secretion assay results are indicative of SEQ ID NO: 113-259 being a diagnostic marker and/or therapeutic target for cancer.

Example 2a: Gene Expression Analysis

Custom Microarray Experiment - Cancer

Custom oligonucleotide microarrays were provided by Agilent Technologies, Inc. (Palo Alto, CA). The microarrays were fabricated by Agilent using their technology for the *in-situ* synthesis of 60mer oligonucleotides (Hughes, et al. 2001, Nature Biotechnology 19:342-347). The 60mer microarray probes were designed by Agilent, from gene

sequences provided by diaDexus, using Agilent proprietary algorithms. Whenever possible two different 60mers were designed for each gene of interest.

All microarray experiments were two-color experiments and were preformed using Agilent-recommended protocols and reagents. Briefly, each microarray was hybridized with cRNAs synthesized from RNA (total RNA for ovarian and prostate, polyA+ RNA for lung, breast and colon samples), isolated from cancer and normal tissues, labeled with fluorescent dyes Cyanine3 (Cy3) or Cyanine5 (Cy5) (NEN Life Science Products, Inc., Boston, MA) using a linear amplification method (Agilent). In each experiment the experimental sample was RNA isolated from cancer tissue from a single individual and the reference sample was a pool of RNA isolated from normal tissues of the same organ as the cancerous tissue (*i.e.* normal ovarian tissue in experiments with ovarian cancer samples). Hybridizations were carried out at 60°C, overnight using Agilent *in-situ* hybridization buffer. Following washing, arrays were scanned with a GenePix 4000B Microarray Scanner (Axon Instruments, Inc., Union City, CA). The resulting images were analyzed with GenePix Pro 3.0 Microarray Acquisition and Analysis Software (Axon).

Data normalization and expression profiling were done with Expressionist software from GeneData Inc. (Daly City, CA/Basel, Switzerland). Gene expression analysis was performed using only experiments that met certain quality criteria. The quality criteria that experiments must meet are a combination of evaluations performed by the Expressionist software and evaluations performed manually using raw and normalized data. To evaluate raw data quality, detection limits (the mean signal for a replicated negative control + 2 Standard Deviations (SD)) for each channel were calculated. The detection limit is a measure of non-specific hybridization. Acceptable detection limits were defined for each dye (<80 for Cy5 and <150 for Cy3). Arrays with poor detection limits in one or both channels were not analyzed and the experiments were repeated. To evaluate normalized data quality, positive control elements included in the array were utilized. These array features should have a mean ratio of 1 (no differential expression). If these features have a mean ratio of greater than 1.5-fold up or down, the experiments were not analyzed further and were repeated. In addition to traditional scatter plots demonstrating the distribution of signal in each experiment, the Expressionist software also has minimum thresholding criteria that employ user defined parameters to identify quality data. These thresholds include two distinct quality measurements: 1) minimum area percentage, which is a measure of the integrity of each spot and 2) signal to noise

ratio, which ensures that the signal being measured is significantly above any background (nonspecific) signal present. Only those features that met the threshold criteria were included in the filtering and analyses carried out by Expressionist. The thresholding settings employed require a minimum area percentage of 60% [(% pixels > background + 2SD)-(% pixels saturated)], and a minimum signal to noise ratio of 2.0 in both channels. By these criteria, very low expressors, saturated features and spots with abnormally high local background were not included in analysis.

Relative expression data was collected from Expressionist based on filtering and clustering analyses. Up-regulated genes were identified using criteria for the percentage of experiments in which the gene is up-regulated by at least 2-fold. In general, up-regulation in ~30% of samples tested was used as a cutoff for filtering.

Two microarray experiments were preformed for each normal and cancer tissue pair. The tissue specific Array Chip for each cancer tissue is a unique microarray specific to that tissue and cancer. The Multi-Cancer Array Chip is a universal microarray that was hybridized with samples from each of the cancers (ovarian, breast, colon, lung, and prostate). See the description below for the experiments specific to the different cancers.

Microarray Experiments and Data Tables

COLON CANCER CHIPS

For colon cancer two different chip designs were evaluated with overlapping sets of a total of 38 samples, comparing the expression patterns of colon cancer derived polyA+ RNA to polyA+ RNA isolated from a pool of 7 normal colon tissues. For the Colon Array Chip all 38 samples (23 Ascending colon carcinomas and 15 Rectosigmoidal carcinomas including: 5 stage I cancers, 15 stage II cancers, 15 stage III and 2 stage IV cancers, as well as 28 Grade1/2 and 10 Grade 3 cancers) were analyzed. The histopathologic grades for cancer are classified as follows: GX, cannot be assessed; G1, well differentiated; G2, Moderately differentiated; G3, poorly differentiated; and G4, undifferentiated. AJCC Cancer Staging Handbook, 5th Edition, 1998, page 9. For the Colon Array Chip analysis, samples were further divided into groups based on the expression pattern of the known colon cancer associated gene Thymidilate Synthase (TS) (13 TS up 25 TS not up). The association of TS with advanced colorectal cancer is well documented. Paradiso *et al.*, *Br J Cancer* 82(3):560-7 (2000); Etienne *et al.*, *J Clin Oncol.* 20(12):2832-43 (2002); Aschele *et al. Clin Cancer Res.* 6(12):4797-802 (2000).

For the Multi-Cancer Array Chip a subset of 27 of these samples (14 Ascending colon carcinomas and 13 Rectosigmoidal carcinomas including: 3 stage I cancers, 9 stage II cancers, 13 stage III and 2 stage IV cancers) were assessed.

The results for the statistically significant up-regulated genes on the Colon Array Chip are shown in Tables 1 and 2. The results for the statistically significant up-regulated genes on the Multi-Cancer Array Chip are shown in Table 3.

The first two columns of each table contain information about the sequence itself (Seq ID, Oligo Name), the next columns show the results obtained for all ("ALL") the colon samples, ascending colon carcinomas ("ASC"), Rectosigmoidal carcinomas ("RS"), cancers corresponding to stages I and II ("ST1,2"), stages III and IV ("ST3,4"), grades 1 and 2 ("GR1,2"), grade 3 ("GR3"), cancers exhibiting up-regulation of the TS gene ("TSup") or those not exhibiting up-regulation of the TS gene ("NOT TSup"). '%up' indicates the percentage of all experiments in which up-regulation of at least 2-fold was observed n=38 for the Colon Array Chip (n=27 for the Multi-Cancer Array Chip), '%valid up' indicates the percentage of experiments with valid expression values in which up-regulation of at least 2-fold was observed.

Table 1.

DEX ID	Oligo Name	Cln ALL %up n=38	Cln ALL % valid up n=38	Cln ASC %up n=23	Cln ASC % valid up n=23	Cln RS %up n=15	Cln RS % valid up n=15	Cln ST1,2 %up n=20	Cln ST1,2 % valid up n=20	Cln ST3,4 %up n=18	Cln ST3,4 %valid up n=18
DEX0449_001.nt.1	10296.0	44.7	44.7	60.9	60.9	20.0	20.0	45.0	45.0	44.4	44.4
DEX0449_001.nt.1	10297.0	63.2	63.2	73.9	73.9	46.7	46.7	65.0	65.0	61.1	61.1
DEX0449_044.nt.1	33264.0	5.3	8.0	0.0	0.0	13.3	25.0	0.0	0.0	11.1	15.4
DEX0449_044.nt.1	33265.0	10.5	11.8	4.3	5.0	20.0	21.4	5.0	5.6	16.7	18.8
DEX0449_044.nt.2	33264.0	5.3	8.0	0.0	0.0	13.3	25.0	0.0	0.0	11.1	15.4
DEX0449_044.nt.2	33265.0	10.5	11.8	4.3	5.0	20.0	21.4	5.0	5.6	16.7	18.8
DEX0449_044.nt.3	33264.0	5.3	8.0	0.0	0.0	13.3	25.0	0.0	0.0	11.1	15.4
DEX0449_044.nt.3	33265.0	10.5	11.8	4.3	5.0	20.0	21.4	5.0	5.6	16.7	18.8
DEX0449_044.nt.4	33264.0	5.3	8.0	0.0	0.0	13.3	25.0	0.0	0.0	11.1	15.4
DEX0449_044.nt.4	33265.0	10.5	11.8	4.3	5.0	20.0	21.4	5.0	5.6	16.7	18.8
DEX0449_046.nt.1	33720.0	36.8	36.8	43.5	43.5	26.7	26.7	35.0	35.0	38.9	38.9

DEX0449_047.nt.1	12809.0	15.8	15.8	13.0	13.0	20.0	20.0	5.0	5.0	27.8	27.8
DEX0449_047.nt.1	12810.0	15.8	15.8	13.0	13.0	20.0	20.0	5.0	5.0	27.8	27.8
DEX0449_047.nt.2	12809.0	15.8	15.8	13.0	13.0	20.0	20.0	5.0	5.0	27.8	27.8
DEX0449_047.nt.2	12810.0	15.8	15.8	13.0	13.0	20.0	20.0	5.0	5.0	27.8	27.8
DEX0449_048.nt.1	28637.0	73.7	73.7	82.6	82.6	60.0	60.0	70.0	70.0	77.8	77.8
DEX0449_048.nt.1	28638.0	65.8	65.8	73.9	73.9	53.3	53.3	70.0	70.0	61.1	61.1
DEX0449_049.nt.1	10224.0	13.2	13.9	17.4	18.2	6.7	7.1	10.0	10.0	16.7	18.8
DEX0449_049.nt.1	10225.0	7.9	8.8	8.7	9.5	6.7	7.7	0.0	0.0	16.7	20.0
DEX0449_051.nt.1	28403.0	10.5	10.5	17.4	17.4	0.0	0.0	10.0	10.0	11.1	11.1
DEX0449_052.nt.1	16385.0	21.1	21.1	21.7	21.7	20.0	20.0	20.0	20.0	22.2	22.2
DEX0449_053.nt.1	17654.0	5.3	5.6	4.3	4.5	6.7	7.1	0.0	0.0	11.1	11.1
DEX0449_053.nt.1	17655.0	13.2	13.9	17.4	19.0	6.7	6.7	10.0	10.5	16.7	17.6
DEX0449_053.nt.2	17654.0	5.3	5.6	4.3	4.5	6.7	7.1	0.0	0.0	11.1	11.1
DEX0449_053.nt.2	17655.0	13.2	13.9	17.4	19.0	6.7	6.7	10.0	10.5	16.7	17.6
DEX0449_054.nt.1	37943.0	52.6	52.6	56.5	56.5	46.7	46.7	50.0	50.0	55.6	55.6
DEX0449_054.nt.1	37944.0	50.0	51.4	56.5	56.5	40.0	42.9	50.0	52.6	50.0	50.0
DEX0449_054.nt.2	37943.0	52.6	52.6	56.5	56.5	46.7	46.7	50.0	50.0	55.6	55.6
DEX0449_054.nt.2	37944.0	50.0	51.4	56.5	56.5	40.0	42.9	50.0	52.6	50.0	50.0
DEX0449_055.nt.1	30951.0	65.8	65.8	69.6	69.6	60.0	60.0	65.0	65.0	66.7	66.7
DEX0449_055.nt.1	30956.0	55.3	56.8	69.6	72.7	33.3	33.3	50.0	52.6	61.1	61.1
DEX0449_055.nt.1	31201.0	65.8	65.8	69.6	69.6	60.0	60.0	70.0	70.0	61.1	61.1
DEX0449_055.nt.2	30951.0	65.8	65.8	69.6	69.6	60.0	60.0	65.0	65.0	66.7	66.7
DEX0449_055.nt.2	30956.0	55.3	56.8	69.6	72.7	33.3	33.3	50.0	52.6	61.1	61.1
DEX0449_055.nt.2	31201.0	65.8	65.8	69.6	69.6	60.0	60.0	70.0	70.0	61.1	61.1
DEX0449_055.nt.3	30951.0	65.8	65.8	69.6	69.6	60.0	60.0	65.0	65.0	66.7	66.7
DEX0449_055.nt.3	30956.0	55.3	56.8	69.6	72.7	33.3	33.3	50.0	52.6	61.1	61.1
DEX0449_055.nt.3	31201.0	65.8	65.8	69.6	69.6	60.0	60.0	70.0	70.0	61.1	61.1

Table 2.

DEX ID	Oligo Name	Cln GR1,2 %up n=28	Cln GR1,2 %valid up n=28	Cln GR3 %up n=10	Cln GR3 %valid up n=10	Cln TS up %up n=13	Cln TS up %valid up n=13	Cln NOT TS up %up n=25	Cln NOT TS up %valid up n=25
DEX0449_001.nt.1	10296.0	35.7	35.7	70.0	70.0	61.5	61.5	36.0	36.0
DEX0449_001.nt.1	10297.0	53.6	53.6	90.0	90.0	84.6	84.6	52.0	52.0
DEX0449_044.nt.1	33264.0	7.1	11.1	0.0	0.0	0.0	0.0	8.0	11.1
DEX0449_044.nt.1	33265.0	10.7	11.5	10.0	12.5	7.7	11.1	12.0	12.0
DEX0449_044.nt.2	33264.0	7.1	11.1	0.0	0.0	0.0	0.0	8.0	11.1
DEX0449_044.nt.2	33265.0	10.7	11.5	10.0	12.5	7.7	11.1	12.0	12.0
DEX0449_044.nt.3	33264.0	7.1	11.1	0.0	0.0	0.0	0.0	8.0	11.1
DEX0449_044.nt.3	33265.0	10.7	11.5	10.0	12.5	7.7	11.1	12.0	12.0
DEX0449_044.nt.4	33264.0	7.1	11.1	0.0	0.0	0.0	0.0	8.0	11.1
DEX0449_044.nt.4	33265.0	10.7	11.5	10.0	12.5	7.7	11.1	12.0	12.0
DEX0449_046.nt.1	33720.0	28.6	28.6	60.0	60.0	38.5	38.5	36.0	36.0
DEX0449_047.nt.1	12809.0	10.7	10.7	30.0	30.0	15.4	15.4	16.0	16.0
DEX0449_047.nt.1	12810.0	10.7	10.7	30.0	30.0	15.4	15.4	16.0	16.0
DEX0449_047.nt.2	12809.0	10.7	10.7	30.0	30.0	15.4	15.4	16.0	16.0
DEX0449_047.nt.2	12810.0	10.7	10.7	30.0	30.0	15.4	15.4	16.0	16.0
DEX0449_048.nt.1	28637.0	71.4	71.4	80.0	60.0	69.2	69.2	76.0	76.0
DEX0449_048.nt.1	28638.0	64.3	64.3	70.0	70.0	61.5	61.5	68.0	68.0
DEX0449_049.nt.1	10224.0	3.6	3.7	40.0	44.4	30.8	30.8	4.0	4.3
DEX0449_049.nt.1	10225.0	3.6	3.8	20.0	25.0	23.1	23.1	0.0	0.0
DEX0449_051.nt.1	28403.0	3.6	3.6	30.0	30.0	23.1	23.1	4.0	4.0
DEX0449_052.nt.1	16385.0	14.3	14.3	40.0	40.0	30.8	30.8	16.0	16.0
DEX0449_053.nt.1	17654.0	0.0	0.0	20.0	20.0	15.4	16.7	0.0	0.0
DEX0449_053.nt.1	17655.0	7.1	7.4	30.0	33.3	38.5	38.5	0.0	0.0
DEX0449_053.nt.2	17654.0	0.0	0.0	20.0	20.0	15.4	16.7	0.0	0.0
DEX0449_053.nt.2	17655.0	7.1	7.4	30.0	33.3	38.5	38.5	0.0	0.0
DEX0449_054.nt.1	37943.0	46.4	46.4	70.0	70.0	76.9	76.9	40.0	40.0
DEX0449_054.nt.1	37944.0	42.9	44.4	70.0	70.0	76.9	76.9	36.0	37.5
DEX0449_054.nt.2	37943.0	46.4	46.4	70.0	70.0	76.9	76.9	40.0	40.0
DEX0449_054.nt.2	37944.0	42.9	44.4	70.0	70.0	76.9	76.9	36.0	37.5
DEX0449_055.nt.1	30951.0	67.9	67.9	60.0	60.0	100.0	100.0	48.0	48.0
DEX0449_055.nt.1	30956.0	53.6	55.6	60.0	60.0	92.3	92.3	36.0	37.5
DEX0449_055.nt.1	31201.0	67.9	67.9	60.0	60.0	100.0	100.0	48.0	48.0
DEX0449_055.nt.2	30951.0	67.9	67.9	60.0	60.0	100.0	100.0	48.0	48.0
DEX0449_055.nt.2	30956.0	53.6	55.6	60.0	60.0	92.3	92.3	36.0	37.5
DEX0449_055.nt.2	31201.0	67.9	67.9	60.0	60.0	100.0	100.0	48.0	48.0
DEX0449_055.nt.3	30951.0	67.9	67.9	60.0	60.0	100.0	100.0	48.0	48.0
DEX0449_055.nt.3	30956.0	53.6	55.6	60.0	60.0	92.3	92.3	36.0	37.5
DEX0449_055.nt.3	31201.0	67.9	67.9	60.0	60.0	100.0	100.0	48.0	48.0

Table 3.

DEX ID	Oligo Name	Cln Multi-Cancer ALL %up n=27	Cln Multi-Cancer ALL %valid up n=27	Cln Multi-Cancer ASC %up n=14	Cln Multi-Cancer ASC %valid up n=14	Cln Multi-Cancer RS %up n=13	Cln Multi-Cancer RS %valid up n=13
DEX0449_044.nt.1	78951.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	78952.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	78952.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	92319.0	0.0	0.0	0.0	0.0	0.0	0.0

DEX0449_044.nt.1	92319.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	92320.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	92320.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	78951.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	78951.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	78952.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	78952.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92319.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92319.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92320.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92320.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	78951.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	78951.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	78952.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	78952.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92319.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92319.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92320.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92320.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	78951.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	78951.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	78952.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	78952.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92319.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92319.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92320.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92320.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_045.nt.1	1045.0	25.9	25.9	14.3	14.3	38.5	38.5

BREAST CANCER CHIPS

For breast cancer two different chip designs were evaluated with overlapping sets of a total of 36 samples, comparing the expression patterns of breast cancer derived polyA+ RNA to polyA+ RNA isolated from a pool of 10 normal breast tissues. For the Breast Array Chip, all 36 samples (9 stage I cancers, 23 stage II cancers, 4 stage III cancers) were analyzed. These samples also represented 10 Grade1/2 and 26 Grade 3 cancers. The histopathologic grades for cancer are classified as follows: GX, cannot be assessed; G1, well differentiated; G2, moderately differentiated; G3, poorly differentiated; and G4, undifferentiated. AJCC Cancer Staging Handbook, pp. 9, (5th Ed, 1998). Samples were further grouped based on the expression patterns of the known breast cancer associated genes Her2 and ERα (10 HER2 up, 26 HER2 not up, 20 ER up and 16 ER not up) and for the Multi-Cancer Array Chip, a subset of 20 of these samples (9 stage I cancers, 8 stage II cancers, 3 stage III cancers) were assessed.

No results for the statistically significant up-regulated genes on the Breast Array Chip are shown. The results for the statistically significant up-regulated genes on the Multi-Cancer Array Chip are shown in Table 4. The first two columns of each table

contain information about the sequence itself (Seq ID, Oligo Name), the next columns show the results obtained for all ("ALL") breast cancer samples, cancers corresponding to stageI ("ST1"), stages II and III ("ST2,3"), grades 1 and 2 ("GR1,2"), grade 3 ("GR3"), cancers exhibiting up-regulation of Her2 ("HER2up") or ERα ("ERup") or those not exhibiting up-regulation of Her2 ("NOT HER2up") or ERα ("NOT ERup"). '%up' indicates the percentage of all experiments in which up-regulation of at least 2-fold was observed (n=36 for Colon Array Chip, n=20 for the Multi-Cancer Array Chip), '%valid up' indicates the percentage of experiments with valid expression values in which up-regulation of at least 2-fold was observed.

Table 4.

DEX ID	Oligo Name	Mam Multi-Cancer ALL %up n=20	Mam Multi-Cancer ALL %valid up n=20	Mam Multi-Cancer ST1 %up n=9	Mam Multi-Cancer ST1 %valid up n=9	Mam Multi-Cancer ST2,3 %up n=11	Mam Multi-Cancer ST2,3 %valid up n=11
DEX0449_044.nt.1	78951.1	5.0	14.3	0.0	0.0	9.1	16.7
DEX0449_044.nt.1	78952.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	78952.1	5.0	14.3	0.0	0.0	9.1	16.7
DEX0449_044.nt.1	92319.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	92319.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	92320.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	92320.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	78951.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	78951.1	5.0	14.3	0.0	0.0	9.1	16.7
DEX0449_044.nt.2	78952.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	78952.1	5.0	14.3	0.0	0.0	9.1	16.7
DEX0449_044.nt.2	92319.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92319.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92320.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92320.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	78951.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	78951.1	5.0	14.3	0.0	0.0	9.1	16.7
DEX0449_044.nt.3	78952.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	78952.1	5.0	14.3	0.0	0.0	9.1	16.7
DEX0449_044.nt.3	92319.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92319.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92320.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92320.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	78951.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	78951.1	5.0	14.3	0.0	0.0	9.1	16.7
DEX0449_044.nt.4	78952.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	78952.1	5.0	14.3	0.0	0.0	9.1	16.7
DEX0449_044.nt.4	92319.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92319.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92320.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92320.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_045.nt.1	1045.0	0.0	0.0	0.0	0.0	0.0	0.0

LUNG CANCER CHIPS

For lung cancer two different chip designs were evaluated with overlapping sets of a total of 29 samples, comparing the expression patterns of lung cancer derived polyA+ RNA to polyA+ RNA isolated from a pool of 12 normal lung tissues. For the Lung Array Chip all 29 samples (15 squamous cell carcinomas and 14 adenocarcinomas including 14 stage I and 15 stage II/III cancers) were analyzed and for the Multi-Cancer Array Chip a subset of 22 of these samples (10 squamous cell carcinomas, 12 adenocarcinomas) were assessed.

The results for the statistically significant up-regulated genes on the Lung Array Chip are shown in Table 5. The results for the statistically significant up-regulated genes on the Multi-Cancer Array Chip are shown in Table 6. The first two columns of each table contain information about the sequence itself (DEX ID, Oligo Name), the next columns show the results obtained for all ("ALL") lung cancer samples, squamous cell carcinomas ("SQ"), adenocarcinomas ("AD"), or cancers corresponding to stage I ("ST1"), or stages II and III ("ST2,3"). '%up' indicates the percentage of all experiments in which up-regulation of at least 2-fold was observed (n=29 for Lung Array Chip, n=22 for Multi-Cancer Array Chip), '%valid up' indicates the percentage of experiments with valid expression values in which up-regulation of at least 2-fold was observed.

Table 5.

DEX ID	Oligo Name	Lng ALL %up n=29	Lng ALL %valid up n=29	Lng SQ %up n=15	Lng SQ %valid up n=15	Lng AD %up n=14	Lng AD %valid up n=14	Lng ST1 %up n=14	Lng ST1 %valid up n=14	Lng ST2,3 %up n=15	Lng ST2,3 %valid up n=15
DEX0449_045 .nt.1	1044.0	13.8	13.8	13.3	13.3	14.3	14.3	14.3	14.3	13.3	13.3
DEX0449_052 .nt.1	5819.0	6.9	7.1	6.7	6.7	7.1	7.7	7.1	7.7	6.7	6.7
DEX0449_052 .nt.1	5820.0	6.9	6.9	6.7	6.7	7.1	7.1	7.1	7.1	6.7	6.7
DEX0449_054 .nt.1	6889.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_054 .nt.1	6890.0	6.9	7.1	6.7	7.1	7.1	7.1	14.3	15.4	0.0	0.0
DEX0449_054 .nt.2	6889.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_054 .nt.2	6890.0	6.9	7.1	6.7	7.1	7.1	7.1	14.3	15.4	0.0	0.0

Table 6.

DEX ID	Oligo Name	Lng Multi-Cancer ALL %up n=22	Lng Multi-Cancer ALL %valid up n=22	Lng Multi-Cancer SQ %up n=10	Lng Multi-Cancer SQ %valid up n=10	Lng Multi-Cancer AD %up n=12	Lng Multi-Cancer AD %valid up n=12
DEX0449_044.nt.1	78951.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	78952.0	4.5	33.3	10.0	50.0	0.0	0.0
DEX0449_044.nt.1	78952.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	92319.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	92319.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	92320.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	92320.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	78951.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	78951.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	78952.0	4.5	33.3	10.0	50.0	0.0	0.0
DEX0449_044.nt.2	78952.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92319.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92319.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92320.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92320.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	78951.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	78951.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	78952.0	4.5	33.3	10.0	50.0	0.0	0.0
DEX0449_044.nt.3	78952.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92319.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92319.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92320.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92320.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	78951.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	78951.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	78952.0	4.5	33.3	10.0	50.0	0.0	0.0
DEX0449_044.nt.4	78952.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92319.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92319.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92320.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92320.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_045.nt.1	1045.0	4.5	4.5	0.0	0.0	8.3	8.3

OVARIAN CANCER CHIPS

For ovarian cancer two different chip designs were evaluated with overlapping sets of a total of 19 samples, comparing the expression patterns of ovarian cancer derived total RNA to total RNA isolated from a pool of 9 normal ovarian tissues. For the Multi-Cancer Array Chip, all 19 samples (14 invasive carcinomas, 5 low malignant potential samples were analyzed and for the Ovarian Array Chip, a subset of 17 of these samples (13 invasive carcinomas, 4 low malignant potential samples) were assessed.

No results for the statistically significant up-regulated genes on the Ovarian Array Chip are shown. The results for the Multi-Cancer Array Chip are shown in Table 7. The first two columns of each table contain information about the sequence itself (DEX ID,

Oligo Name), the next columns show the results obtained for all ("ALL") ovarian cancer samples, invasive carcinomas ("INV") and low malignant potential ("LMP") samples. '%up' indicates the percentage of all experiments in which up-regulation of at least 2-fold was observed (n=19 for the Multi-Cancer Array Chip, n=17 for the Ovarian Array Chip), '%valid up' indicates the percentage of experiments with valid expression values in which up-regulation of at least 2-fold was observed.

Table 7.

DEX ID	Oligo Name	Ovr Multi-Cancer ALL %up n=19	Ovr Multi-Cancer ALL %valid up n=19	Ovr Multi-Cancer INV %up n=14	Ovr Multi-Cancer INV %valid up n=14	Ovr Multi-Cancer LMP %up n=5	Ovr Multi-Cancer LMP %valid up n=5
DEX0449_044.nt.1	78951.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	78952.0	5.3	12.5	7.1	16.7	0.0	0.0
DEX0449_044.nt.1	78952.1	5.3	12.5	7.1	14.3	0.0	0.0
DEX0449_044.nt.1	92319.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	92319.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	92320.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	92320.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	78951.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	78951.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	78952.0	5.3	12.5	7.1	16.7	0.0	0.0
DEX0449_044.nt.2	78952.1	5.3	12.5	7.1	14.3	0.0	0.0
DEX0449_044.nt.2	92319.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92319.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92320.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92320.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	78951.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	78951.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	78952.0	5.3	12.5	7.1	16.7	0.0	0.0
DEX0449_044.nt.3	78952.1	5.3	12.5	7.1	14.3	0.0	0.0
DEX0449_044.nt.3	92319.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92319.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92320.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92320.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	78951.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	78951.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	78952.0	5.3	12.5	7.1	16.7	0.0	0.0
DEX0449_044.nt.4	78952.1	5.3	12.5	7.1	14.3	0.0	0.0
DEX0449_044.nt.4	92319.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92319.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92320.0	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92320.1	0.0	0.0	0.0	0.0	0.0	0.0
DEX0449_045.nt.1	1045.0	0.0	0.0	0.0	0.0	0.0	0.0

PROSTATE CANCER

For prostate cancer three different chip designs were evaluated with overlapping sets of a total of 29 samples, comparing the expression patterns of prostate cancer or

benign disease derived total RNA to total RNA isolated from a pool of 35 normal prostate tissues. For the Prostate 1 Array and Prostate 2 Array Chips all 29 samples (17 prostate cancer samples, 12 non-malignant disease samples) were analyzed. For the Multi-Cancer Array Chip a subset of 28 of these samples (16 prostate cancer samples, 12 non-malignant disease samples) were analyzed.

The results for the statistically significant up-regulated genes on the Prostate1 Array Chip and the Prostate2 Array Chip are shown in Table 8. The results for the statistically significant up-regulated genes on the Multi-Cancer Array Chip are shown in Table 9. The first two columns of each table contain information about the sequence itself (DEX ID, Oligo Name), the next columns show the results obtained for prostate cancer samples ("CAN") or non-malignant disease samples ("DIS"). '%up' indicates the percentage of all experiments in which up-regulation of at least 2-fold was observed (n=29 for the Prostate1 Array Chip and the Multi-Cancer Array Chip), '%valid up' indicates the percentage of experiments with valid expression values in which up-regulation of at least 2-fold was observed.

Table 8.

DEX ID	Oligo Name	Pro CAN %up n=17	Pro CAN %valid up n=17	Pro DIS %up n=12	Pro DIS %valid up n=12
DEX0449_045.nt.1	27017.01	0.0	0.0	0.0	0.0
DEX0449_045.nt.1	27017.02	0.0	0.0	0.0	0.0
DEX0449_045.nt.1	28249.01	0.0	0.0	0.0	0.0
DEX0449_045.nt.1	28249.02	0.0	0.0	0.0	0.0
DEX0449_045.nt.3	27017.01	0.0	0.0	0.0	0.0
DEX0449_045.nt.3	27017.02	0.0	0.0	0.0	0.0
DEX0449_047.nt.1	31094.01	0.0	0.0	0.0	0.0
DEX0449_047.nt.1	31094.02	0.0	0.0	0.0	0.0
DEX0449_047.nt.1	31094.03	0.0	0.0	0.0	0.0
DEX0449_047.nt.1	33588.01	5.9	8.3	0.0	0.0
DEX0449_047.nt.1	33588.02	0.0	0.0	0.0	0.0
DEX0449_047.nt.2	31094.01	0.0	0.0	0.0	0.0
DEX0449_047.nt.2	31094.02	0.0	0.0	0.0	0.0
DEX0449_047.nt.2	31094.03	0.0	0.0	0.0	0.0
DEX0449_047.nt.2	33588.01	5.9	8.3	0.0	0.0
DEX0449_047.nt.2	33588.02	0.0	0.0	0.0	0.0
DEX0449_051.nt.1	35746.01	0.0	0.0	0.0	0.0
DEX0449_051.nt.1	35746.02	0.0	0.0	0.0	0.0
DEX0449_051.nt.1	35746.03	0.0	0.0	0.0	0.0
DEX0449_051.nt.1	35788.01	0.0	0.0	0.0	0.0
DEX0449_051.nt.1	35788.02	0.0	0.0	0.0	0.0
DEX0449_051.nt.1	35788.03	0.0	0.0	0.0	0.0

Table 9.

DEX ID	Oligo Name	Pro Multi-Cancer CAN %up n=16	Pro Multi-Cancer CAN %valid up n=16	Pro Multi-Cancer DIS %up n=12	Pro Multi-Cancer DIS %valid up n=12
DEX0449_044.nt.1	78951.1	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	78952.0	5.9	7.7	0.0	0.0
DEX0449_044.nt.1	78952.1	5.9	7.1	0.0	0.0
DEX0449_044.nt.1	92319.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	92319.1	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	92320.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.1	92320.1	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	78951.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	78951.1	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	78952.0	5.9	7.7	0.0	0.0
DEX0449_044.nt.2	78952.1	5.9	7.1	0.0	0.0
DEX0449_044.nt.2	92319.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92319.1	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92320.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.2	92320.1	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	78951.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	78951.1	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	78952.0	5.9	7.7	0.0	0.0
DEX0449_044.nt.3	78952.1	5.9	7.1	0.0	0.0
DEX0449_044.nt.3	92319.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92319.1	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92320.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.3	92320.1	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	78951.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	78951.1	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	78952.0	5.9	7.7	0.0	0.0
DEX0449_044.nt.4	78952.1	5.9	7.1	0.0	0.0
DEX0449_044.nt.4	92319.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92319.1	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92320.0	0.0	0.0	0.0	0.0
DEX0449_044.nt.4	92320.1	0.0	0.0	0.0	0.0
DEX0449_045.nt.1	1045.0	0.0	0.0	0.0	0.0

SEQ ID NO: 1-112 was up-regulated on various tissue microarrays. Accordingly, nucleotide SEQ ID NO: 1-112 or the encoded protein SEQ ID NO: 113-259 may be used as a cancer therapeutic and/or diagnostic target for the tissues in which expression is shown.

The following table lists the location (Oligo Location) where the microarray oligos (Oligo ID) map on the transcripts (DEX ID) of the present invention. Each Oligo ID may have been printed multiple times on a single chip as replicates. The Oligo Name is an exemplary replicate (e.g. 1000.01) for the Oligo ID (e.g. 1000), and data from other replicates (e.g. 1000.02, 1000.03) may be reported. Additionally, the Array (Chip Name) that each oligo and oligo replicates were printed on is included.

DEX NT ID	Oligo ID	Oligo Name	Chip Name	Oligo Location
DEX0449_001.nt.1	10296	10296.0	Colon array	1034-1093
DEX0449_001.nt.1	10297	10297.0	Colon array	287-346
DEX0449_044.nt.1	78952	78952.0	Multi-Cancer array	428-487
DEX0449_044.nt.1	33264	33264.0	Colon array	405-464
DEX0449_044.nt.1	92319	92319.0	Multi-Cancer array	546-605
DEX0449_044.nt.1	78951	78951.0	Multi-Cancer array	425-484
DEX0449_044.nt.1	33265	33265.0	Colon array	356-415
DEX0449_044.nt.1	92320	92320.0	Multi-Cancer array	526-585
DEX0449_044.nt.2	92319	92319.0	Multi-Cancer array	592-651
DEX0449_044.nt.2	92320	92320.0	Multi-Cancer array	572-631
DEX0449_044.nt.2	33264	33264.0	Colon array	451-510
DEX0449_044.nt.2	78952	78952.0	Multi-Cancer array	474-533
DEX0449_044.nt.2	33265	33265.0	Colon array	402-461
DEX0449_044.nt.2	78951	78951.0	Multi-Cancer array	471-530
DEX0449_044.nt.3	92320	92320.0	Multi-Cancer array	359-418
DEX0449_044.nt.3	33265	33265.0	Colon array	189-248
DEX0449_044.nt.3	78952	78952.0	Multi-Cancer array	261-320
DEX0449_044.nt.3	33264	33264.0	Colon array	238-297
DEX0449_044.nt.3	78951	78951.0	Multi-Cancer array	258-317
DEX0449_044.nt.3	92319	92319.0	Multi-Cancer array	379-438
DEX0449_044.nt.4	92320	92320.0	Multi-Cancer array	335-394
DEX0449_044.nt.4	78952	78952.0	Multi-Cancer array	237-296
DEX0449_044.nt.4	33265	33265.0	Colon array	165-224
DEX0449_044.nt.4	92319	92319.0	Multi-Cancer array	355-414
DEX0449_044.nt.4	33264	33264.0	Colon array	214-273
DEX0449_044.nt.4	78951	78951.0	Multi-Cancer array	234-293
DEX0449_045.nt.1	1044	1044.0	Lung array	175-234
DEX0449_045.nt.1	27017	27017.01	Prostate1 array	1043-1102
DEX0449_045.nt.1	1045	1045.0	Multi-Cancer array	165-224
DEX0449_045.nt.1	28249	28249.02	Prostate1 array	323-382
DEX0449_045.nt.2	1044	1044.0	Lung array	175-234
DEX0449_045.nt.2	1045	1045.0	Multi-Cancer array	165-224
DEX0449_045.nt.3	1044	1044.0	Lung array	160-219
DEX0449_045.nt.3	27017	27017.01	Prostate1 array	1613-1672
DEX0449_045.nt.3	1045	1045.0	Multi-Cancer array	150-209
DEX0449_046.nt.1	33720	33720.0	Colon array	1037-1096
DEX0449_047.nt.1	12809	12809.0	Colon array	2811-2870
DEX0449_047.nt.1	33588	33588.01	Prostate1 array	2797-2856
DEX0449_047.nt.1	12810	12810.0	Colon array	2773-2832
DEX0449_047.nt.1	31094	31094.02	Prostate2 array	2319-2378
DEX0449_047.nt.2	12810	12810.0	Colon array	2773-2832
DEX0449_047.nt.2	33588	33588.01	Prostate1 array	2797-2856
DEX0449_047.nt.2	31094	31094.02	Prostate2 array	2319-2378
DEX0449_047.nt.2	12809	12809.0	Colon array	2811-2870
DEX0449_048.nt.1	28637	28637.0	Colon array	357-416
DEX0449_048.nt.1	28638	28638.0	Colon array	216-275
DEX0449_049.nt.1	10224	10224.0	Colon array	360-419
DEX0449_049.nt.1	10225	10225.0	Colon array	242-301
DEX0449_051.nt.1	35788	35788.03	Prostate2 array	1107-1166
DEX0449_051.nt.1	28403	28403.0	Colon array	1107-1166
DEX0449_051.nt.1	35746	35746.02	Prostate2 array	893-952
DEX0449_052.nt.1	5820	5820.0	Lung array	1714-1773
DEX0449_052.nt.1	16385	16385.0	Colon array	2216-2275
DEX0449_052.nt.1	5819	5819.0	Lung array	1865-1924
DEX0449_053.nt.1	17655	17655.0	Colon array	1536-1595

DEX0449_053.nt.1	17654	17654.0	Colon array	1600-1659
DEX0449_053.nt.2	17655	17655.0	Colon array	1055-1114
DEX0449_053.nt.2	17654	17654.0	Colon array	1119-1178
DEX0449_054.nt.1	37944	37944.0	Colon array	1296-1355
DEX0449_054.nt.1	6889	6889.0	Lung array	3086-3145
DEX0449_054.nt.1	6890	6890.0	Lung array	2956-3015
DEX0449_054.nt.1	37943	37943.0	Colon array	1416-1475
DEX0449_054.nt.2	6889	6889.0	Lung array	2461-2520
DEX0449_054.nt.2	37944	37944.0	Colon array	671-730
DEX0449_054.nt.2	6890	6890.0	Lung array	2331-2390
DEX0449_055.nt.1	30956	30956.0	Colon array	4398-4457
DEX0449_055.nt.1	31201	31201.0	Colon array	4333-4392
DEX0449_055.nt.1	30951	30951.0	Colon array	4334-4393
DEX0449_055.nt.2	30956	30956.0	Colon array	1809-1868
DEX0449_055.nt.2	31201	31201.0	Colon array	1744-1803
DEX0449_055.nt.3	30951	30951.0	Colon array	3899-3958
DEX0449_055.nt.3	30956	30956.0	Colon array	3963-4022

Example 2b: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman® probes is a quantitation detection system utilizing the 5'- 3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman®) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman® probes specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are

relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the CSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to the calibrator. Normal RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the CSNA in pairs of matched samples may also be determined. A matched pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual. All the values are compared to the calibrator.

In the analysis of matching samples, the CSNAs show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples. Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer state (*e.g.* higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Informaton on the samples tested in the QPCR experiments include the Sample ID (Smpl ID), Organ, Tissue Type (Tiss Type), Diagnosis (DIAG), Disease Detail, and Stage or Grade (STG or GRD).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matched samples tested are indicative of SEQ ID NO: 1-112 being a diagnostic marker and/or a therapeutic target for cancer.

Example 3: Protein Expression

The CSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the CSNA is subcloned in pET-21d for expression in E. coli. In addition to the CSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH_2-terminus of the coding sequence of CSNA, and six histidines, flanking the COOH-terminus of the coding sequence of CSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of CSP is achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that are separated from total cell lysate were incubated with a nickel chelating resin. The column is packed and washed with five column volumes of wash buffer. CSP is eluted stepwise with various concentration imidazole buffers.

Example 4: Fusion Proteins

The human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5'and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e.g.*, WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from

the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1-112. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al.*, *Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al.*, *Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons are also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al., Nucleic Acids Res.*, 19: 1156

(1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Manheim), and FISH is performed as described in Johnson *et al., Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. Johnson (1991). Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 µl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 µl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1, μg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μg/kg/hour to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), bucally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481, the contents of which are hereby incorporated by reference herein in their entirety), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324, the contents of which are hereby incorporated by reference herein in their entirety. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably, the carrier is a parenteral carrier, more preferably, a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense or RNAi technology are used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5'and 3'end sequences respectively as set forth in Example 3. Preferably, the 5'primer contains an EcoRI site and the 3'primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 12: Method of Treatment Using Gene Therapy-In Vivo

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, Tabata H. *et al. Cardiovasc. Res.* 35 (3): 470-479 (1997); Chao J *et al.* Pharmacol. Res. 35 (6): 517-522 (1997); Wolff J. A. *Neuromuscul. Disord.* 7 (5): 314-318 (1997), Schwartz B. *et al. Gene Ther.* 3 (5): 405-411 (1996); and Tsurumi Y. *et al. Circulation* 94 (12): 3281-3290 (1996); W0 90/11092, W0 98/11779; U. S. Patent No. 5,693,622; 5,705,151; 5,580,859, the contents of which are hereby incorporated by reference herein in their entirety.

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, colon, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. *et al. Ann. NY Acad. Sci.* 772: 126-139 (1995) and Abdallah B. *et al. Biol. Cell* 85 (1): 1-7 (1995)) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the

transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, colon, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. In vivo muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 μg/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to colons or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle in vivo is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e.g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (I. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection

(Paterson et al., *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver et al., *Biotechnology* 11: 1263-1270 (1993); Wright et al., *Biotechnology* 9: 830-834 (1991); and U. S. Pat. No. 4,873,191, the contents of which is hereby incorporated by reference herein in its entirety); retrovirus mediated gene transfer into germ lines (Van der Putten et al., *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, *Mol Cell. Biol.* 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., *Science* 259: 1745 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., *Cell* 57: 717-723 (1989). For a review of such techniques, see Gordon, "Transgenic Animals," *Intl. Rev. Cytol.* 115: 171-229 (1989).

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell *et al., Nature* 380: 64-66 (1996); Wilmut *et al., Nature* 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e.g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., *Science* 265: 103-106 (1994)). The regulatory sequences required for such a

cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503512 (1987);

Thompson et al., Cell 5: 313-321 (1989)) Alternatively, RNAi technology may be used. For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention in vivo. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, supra). However, this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site in vivo using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient in vivo. Such cells may be obtained from the patient (i.e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered in vitro using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e.g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent No. 5,399,349; and Mulligan & Wilson, U. S. Patent No. 5,460,959, the contents of which are hereby incorporated by reference herein in their entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.